The foundations of lesion-function inference in the human brain

Dr Yee-Haur Mah

Institute of Neurology

UCL

Submitted in partial fulfilment of the requirement for a degree of Doctor of Philosophy

I, Yee-Haur Mah confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Understanding the functional architecture of the brain has long been a challenge in neuroscience with a variety of techniques having been developed to explore this structure-function relationship. However, in order to be able to accurately identify the underlying system we require techniques that have the capabilities of describing the complexities therein.

In order to perform lesion-function studies a cohort of brain scans with the location of the lesion identified must be collected. Utilising diffusion weighted magnetic resonance imaging, normally collected in the clinical setting, I propose a new unsupervised lesion segmentation routine.

The cohort of brain scans also need to be spatially normalised such that homologous regions of the brain are brought into register with each other. However, this process can be perturbed by the presence of a lesion within the scan. Though a series of simulations I evaluate the performance of 12 different spatial normalisation routines on brains scans that possess a lesion.

Historically lesion-function mapping studies have tended to use a univariate statistical approach, where different locations within the brain are treated as being spatially independent from each other. Here I show that biases within the structure of the data have the potential to distort the lesion-function inferences we draw. Though a series of simulations, I show that a mass univariate technique is vulnerable to these biases and assess three different multivariate methods (Support Vector Machines, Relevance Vector Machines and Flexible Bayesian Modelling) as potential solutions to this problem.

Asides from making lesion-function inferences, these multivariate models can be used to predict future events. Using a data set of paired admission diffusion

weighted magnetic resonance imaging scans and functional outcome scores I apply these techniques to the clinical scenario of predicting the functional outcome of patients after a cerebral vascular event.

Acknowledgements

I am very grateful to Professor Husain and Dr Nachev for supervising this thesis. Their patience and professionalism throughout this process are deeply appreciated.

I am heavily indebted to Dr Nachev, for introducing me to this field of research. Without his generosity and guidance, none of this work would have been possible, and it is to whom I owe my utmost gratitude.

I would also like to extend my thanks to Professor Jackson (Nottingham University), Dr Yu, So Young Kim and Sun Young Choi (Korea University, Anam Hospital) who kindly assisted in obtaining the data used in §6 of this thesis.

Contents

List of Figures

3 Optimal inter-subject registration of human magnetic resonance brain imaging in the presence of focal lesions

4 Lesion function mapping using mass univariate techniques

6 Post stroke outcome prediction using acute stroke imaging and high-dimensional multivariate algorithms

7 Conclusion

8 Appendix

1 The foundations of lesion-function inference in the human brain

1.1 Introduction

Lesion-function inference refers to understanding the localisation of function in the human brain with the aid of lesions. Over the history of neuroscience a variety of techniques has contributed to our current understanding of this functional architecture. The first of these, historically, were lesion studies that examined the relationship between regions of localised injury and the observed behaviour or functional deficits exhibited by the afflicted patients. In the past, lesion studies relied on post mortem examination of the patient's brain to identify the anatomy of the lesion (Broca, 1861; Wernicke, 1874). However, with the development of non-invasive brain imaging such as computer tomography (CT) and magnetic resonance imaging (MRI), the regions of damage could be examined *in vivo*, therefore not only improving the spatial resolution of the technique but also facilitating the collection of suitable control subjects (Bates et al., 2003; Bird et al., 2006; Rorden and Karnath, 2004). In parallel with this, other techniques for functional brain mapping such as transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS) and especially functional magnetic resonance imaging (fMRI) have emerged: as a result it would be an understatement to say that lesion studies have faded in popularity, particularly relative to fMRI. The question I wish to examine here is whether or

not, and to what extent, lesion mapping remains a valuable tool for determining the functional architecture of the brain.

Functional magnetic resonance imaging (fMRI) has become a popular tool for investigating the functional architecture of the brain as it provides a non-invasive approach to visualising, *in vivo*, the functioning brain (Heeger and Ress, 2002; Matthews and Jezzard, 2004). The technique relies on the different magnetic properties of oxygenated and deoxygenated haemoglobin – blood oxygen level dependent (BOLD) signal – and the association between increased neuronal activity and local changes in oxygen saturation. Critically, the BOLD signal is only an indirect measure of neuronal activity and may be influenced by many factors not captured in the experimental design. In addition, the signal measured appears to be related principally to input, rather than output, from the neural tissue (Logothetis, 2003), reflecting only a limited aspect of neural function. In comparison with lesion studies where the experimenter is confident that the injured areas are no longer functional, the degree of certainty with fMRI with regards to the level of involvement of an identified region is far more variable. Most importantly, even where BOLD is a reliable indicator of neural function the presence or absence of activity can only be *correlated* with a function of interest. A correlation establishes neither necessity nor sufficiency for any function: the two conditions that need to be satisfied for any substrate to be argued to *mediate* a given function. As a result, fMRI is useful in identifying *candidate* regions potentially critical for a function, i.e. generating *hypothetical* functional models of the brain, but less useful in discriminating between them. Logically, in order to test whether a specific region is critical for a particular function, one must be able to show the *loss* of function (or expression of the symptom) when the region is inactivated (Aue et

al., 2009). For strong inference we must therefore turn to techniques that either generate or exploit disruption of brain function.

Transcranial magnetic stimulation (TMS) is a non invasive procedure that renders a region of the brain dysfunctional by the application of a rapidly changing magnetic field over the exterior surface of the skull (Walsh and Cowey, 2000). The fluctuating magnetic field causes the depolarization and hyperpolarizaton of the neurons directly beneath the electric coil, temporarily disturbing the normal functioning of the region of brain. The ability to examine the same brain in the two different states significantly increases the inferential power of the technique as it allows within-subject comparisons, although it should be appreciated that the reported behavioural effects of TMS have largely been on increasing reaction time, rather than altering other aspects of performance (e.g. errors). However, the anatomical range of TMS is limited, particularly in depth, as the effects of stimulation are confined to superficial cortical regions and cannot be used to investigate deep medial or subcortical structures (Epstein et al., 1990; Rudiak and Marg, 1994; Walsh and Cowey, 2000; Zangen et al., 2005). Furthermore, if we were to treat the surface of the brain as a two dimensional surface, the effect of TMS within this plane would always be ill-defined, as the extent of influence exerted by a magnetic field decreases logarithmically with increasing distance from the source (Sack and Linden, 2003).

Procedural modifications have been developed to counteract this issue, include stimulating multiple overlapping regions and determining the localization of the behaviour by subtractive inference, thereby improving the spatial resolution to the order of millimetres. However, it should be borne in mind, that the surface of the brain consists of multiple folds and overlapping gyri, greatly complicating the modelling of field spread. Furthermore, although a specific region of the cortex is stimulated by the TMS coil, the relation of this region to the functional

outcome may not be direct. Remote cortical and subcortical regions may be implicated in generating the observed outcome in a complex network via a series of connections, impacting on the spatial specificity of the technique (Paus et al., 1997; Ruff et al., 2009). Although multiple stimulation sites may be used to help improve the spatial resolution, there is still a practical limitation to the total number of sites one can stimulate concurrently. This will also place a restriction on our ability to test multiple site localisation.

Similar difficulties complicate transcranial direct current stimulation (tDCS), which involves applying an electric current between two electrodes over the surface of the skull to modulate the brain tissue between them (Utz et al., 2010). Unlike TMS, it does not induce polarization and hyperpolarization of the underlying neurons, but modulates their resting potential, thereby modifying the excitability of the neurons. A further difficulty here is the uncertainty about the location of these effects: *any* region of the brain between the electrodes may be affected. As a tool for making anatomical inferences about brain function its power is necessarily limited.

These objections are not exhaustive. But they give us enough grounds to return to lesion-function mapping as a technique that can provide *both* anatomical precision and strong inferential power to test the many hypotheses generated by functional imaging. Though lesions in the human cannot be induced experimentally (for obvious ethical reasons) there is a wealth of such data in clinical populations, in particular from cerebrovascular injury. We need to consider how such data should best be used to make lesion-function inferences in the human brain.

1.2 Overview

In order to make inferences that may be generalised to the population, studies involving cohorts of multiple patients are required (Rorden and Karnath, 2004). As a consequence, unlike single case reports, the neuro-imaging data must be appropriately preprocessed into a format suitable for analysis across a group. The journey from data to inference can be divided conceptually into 3 broad steps. First, the lesioned regions of the brain must be differentiated from the healthy parts. This process of identifying lesioned from healthy tissue is known as *lesion segmentation* (Fiez et al., 2000). Second, the anatomical labels of the brain must be identified. This is achieved by bringing the imaging data into register with a standard labelled template through a process also known as *spatial normalisation* (Ashburner and Friston, 1997, 1999; Friston et al., 1995). This ensures homologous parts of the brain are aligned and in register with a known map, so as to allow across-subject comparisons. Third are the statistical calculations that are performed on the preprocessed data: the *lesion-function inference* proper (Rorden and Karnath, 2004).

Though these are, conceptually speaking, simple practicalities it should be borne in mind that the context is complex. *Any* study here is dependent on the anatomical consistencies – across subjects – of the functional specialization within the brain. If no relationship – *consistent amongst different people* – exists, no technique comparing groups will be able to find one. The preprocessing steps aim to minimize the noise introduced into the process and thus to improve our ability to identify true relationships: but that is as far as they go. Furthermore, the structure-function relationships our inferential techniques use to model the brain must be sensitive to the potential relationships that may exist. In the syndrome of visuospatial neglect (in which patients fail to pay attention towards contralesional space), for example, the effects of a symptomatic lesion in the right parietal lobe can be reversed by

a *second* lesion in the left frontal lobe (Vuilleumier et al., 1996). This would suggest that the brain is not limited to simple monotonic relationships, but has the potential for far more complex, non-monotonic interactions within distributed systems. Our inferential techniques used to model the brain must be able to handle these possibilities. Critically, any bias introduced into the system, such as through the sampling of the lesions, is of grave concern as these effects unlike noise will persist or even amplify as the data set is increased.

1.3 Lesion segmentation

Lesion segmentation is the process of differentiating between healthy and injured tissue, and is essential before any spatial inference about the behavioural effects of a lesion can be made. Inevitably, in clinical practice it reduces to the question of normal vs abnormal brain image *signal*, for we do not ordinarily have any other means of exploring the human brain *in vivo* (Bhanu Prakash et al., 2008; Fiez et al., 2000). The difficulty is that it is not just the signal abnormality at a given point, but the pattern across the brain, that would naturally compel us to label an area as damaged. For example, an island of normal tissue within a damaged area might have entirely normal signal but clearly cannot be functional, for it will be disconnected from the rest of the brain. This kind of complexity, (quite apart from the difficulty of analysing brain images even in terms of isolated signal abnormalities), means the gold standard is manual segmentation: doing it by eye (Mort et al., 2003). With this in mind, it is clear that the process of segmentation and registration are closely intertwined, since deciding whether an area is injured or abnormal relies on comparing its appearance with what we would normally expect to find *at that location*. There is also the issue of the definition of an abnormal signal itself being contextual, and not an absolute value. Not only do brains differ from one person to the next in the normal population, but from a technological stand point it is particularly relevant with magnetic resonance imaging, as there is no standard signal scale.

1.3.1 Manual segmentation

Manual segmentation refers to the process of lesion segmentation performed by a person. A trained operator draws around a lesion on the basis of his expertise, whatever it is that might be said to consist in. The obvious difficulty of not having a perspicuous set of criteria aside, this approach is

labour intensive, rendering large scale analyses infeasible. Moreover, the flexibility that gives manual segmentation its strength is also the source of its greatest weakness: the structure of any bias in drawing around the lesion is unpredictable (itself being very complex), and the variability shows the potential for bias is not insubstantial (Van Leemput et al., 2001; Tan et al., 2002). Indeed, one study examining intra and inter- operator variability on a small set of 10 brains, defined as the percentage of non-overlapping voxels between the binary masks derived from the same lesioned brain image, found a mean variability of 31% and 33% respectively. This is certainly potentially large enough — if associated with bias — to distort data (Fiez et al., 2000).

1.3.2 Automated segmentation

The case in favour of avoiding manual segmentation can be made with ease; the question is what should take its place. We now look at various approaches, in order of decreasing requirement for human input (supervision).

1.3.2.1 Supervised (semi-automated) segmentation

Supervised (semi-automated) segmentation is a very broad category, with a range of computer assistance offered to the operator. Methods to simplify the more complex task of delineating the normal-abnormal boundary have been attempted with algorithms that work “on-the-fly”, guided by the operator’s hand (Barrett and Mortensen, 1997; Falcao et al., 1998, 2000). Whilst the operator assesses the image and identifies “seeds” along the boundary line, the algorithm uses these “seeds” to calculate its own boundary line based on local information, adjusting in real-time the position of the boundary line placement. The operator can then, if necessary, review and adjust sections of the boundary at will. One such method, the ultra fast user-steered “live wire on-the-fly”, allows the operator to visualize the position of the boundary line

calculated as he moved the cursor around the 2d image (Falcao et al., 2000). This arrangement resulted in an increase in operator accuracy, a reduction in variability and a significant drop in time costs of the magnitude of 1.2—31 times.

Alternative supervised methodologies have looked at further reducing operator involvement by trying to completely automate boundary extraction. One such example is by Filippi et al. (Filippi et al., 1995). Here they looked at the effect of (a relatively simplistic method of) computer assisted segmentation on intra- and inter-operator variability. Assistance took the form of a rudimentary thresholding method which subsequently required manual post processing to refine the lesion edge. Just as Fiez et al found, there was indeed intra- and inter-operator variation, but more importantly they reported that their computer assisted method reduced both these forms of variation. Unlike the earlier described method, the first-pass boundary extraction has been performed largely by the algorithm alone, with refinements performed on a second pass analysis.

More recently Ashton et al. (Ashton et al., 2003) examined the performance of 2 semi-automated techniques against manual segmentation, looking specifically at the inherent variability of each method. The first of these semi-automated methods, geometrically constrained region growth (GEORG), required the operator simply to identify each of the lesions present in the brain volume with a single mouse click. The algorithm would subsequently identify the boundaries of each of the lesions flagged by the operator. In the second, directed multispectral segmentation (DMSS), an example lesion within the brain volume needed to be first delineated by the operator (either manually or semi-automated), with the algorithm then searching the rest of the brain volume for further lesions. This method proved itself to be very useful to screen for any missed lesions, particularly in diseases like multiple sclerosis, where

there are usually multiple lesions, but was less helpful in diseases generally characterized by a single lesion. Nevertheless, manual segmentation was shown to exhibit the most amount of intra- and inter-operator variability, with both semi-automated algorithms displaying improved accuracy and reduced variability on artificially lesioned brain volumes. The authors suggested that some of the variability was due to the remaining operator involvement as the DMSS algorithm is dependent on the lesion selected and the GEORG method required the operator identifying each lesion within the brain.

It may be paradoxical that these studies focus on measures of operator variability while retaining the one element sure to guarantee it: the involvement of *any* kind of operator. Despite the improvements seen, a completely automated procedure is clearly the only answer to the problem of unknown bias from an operator. It is to these algorithms that I will dedicate greater space, looking at the various methodologies currently available to us and later examining how they have been adapted to the problem of lesion segmentation.

1.3.2.2 Unsupervised (fully automated) segmentation

The process of labelling an image without the aid of an operator — whether in terms of diseased vs normal or not — is termed unsupervised segmentation. Since there is nothing or little but the data to guide the labelling process, its theoretical background falls within the domains of supervised and unsupervised learning algorithms. With advancements in digital imaging and computing power following Moore's law (where the number of transistors on an integrated circuit doubles approximately every 2 years), there has been increasing interest in developing computer algorithms to fully automate the process. Irrespective of current hardware / technological limitations, designing a fully automated algorithm is still not an easy task. The problems with these algorithms reduce to 3 main points: first, there is no clear definition of what

is correct and what is not. Unless a “ground-truth” answer is available, it is difficult to differentiate between the appropriateness of one result compared with another, particularly with high dimensional data. The feature(s) one algorithm prioritises over another is often unknown to the operator and deciding which of these (abstract) measures is more pertinent is open to debate. All that is amenable to comparison is the final result. Second, these algorithms are very sensitive to the dimensionality of the data set. As the number of dimensions increase, the algorithm must search a more complex feature space, thereby increasing the computational load, generally in a non-linear fashion, and reducing the probability a good solution may be found. Lastly, these algorithms are sensitive to the parameters of the process. A fully automated method is akin to a ballistic motor action. The operator must first set the parameters of the algorithm prior to its execution, and then await the final output result. Although not all these parameters are problematic to identify, some pose significant obstacles to complete automation. For example returning to the semi-automated technique of GEORG the algorithm needs to know *a priori* the number of lesions present in the data set (Ashton et al., 2003).

There is an entire mathematical field dedicated to learning algorithms and I do not claim to provide a comprehensive review (Duda and Hart, 1973; Forgy, 1965; MacKay, 2003; MacQueen, 1967). The following sections aim to provide an introduction to some of the concepts behind the mathematics and how their restrictions may impinge on their applicability to the task of lesion segmentation.

1.3.3 Learning algorithms

To differentiate between healthy and damaged tissue, segmentation algorithms must learn some feature(s) that separates them. Although a complete profile of differences would be optimal, it is not necessary and often not possible

with the data available. Careful consideration is required to appropriately parameterise the data to facilitate the identification of these discriminatory features. Broadly speaking, learning algorithms can be divided into two groups – unsupervised and supervised algorithms – depending on the source of the information on which the discrimination is made.

1.3.3.1 Unsupervised learning algorithms

Unsupervised techniques are methods that do not require *a priori* knowledge of the sub groups within the data, using only the information contained within the test data itself. They are also referred as *clustering* algorithms, identifying structure within the data without the aid of a separate label. Though many such algorithms have been proposed, here I examine two that have found application within our field of interest: *k*-means clustering (Forgy, 1965; MacQueen, 1967) and mean-shift clustering (Fukunage and Narendra, 1975).

1.3.3.1.1 *k*-means clustering

K-means clustering is an iterative algorithm that seeks to partition a dataset into a fixed number (*k*) of groups and achieves this by minimizing some measure of within-group dissimilarity (Forgy, 1965; MacQueen, 1967). By knowing the number of centres within the data set, the algorithm seeks to minimize the maximum distance of every point from its closest centre. Data points are then grouped into their most appropriate cluster where the objective is to minimize the sum distance from its centre.

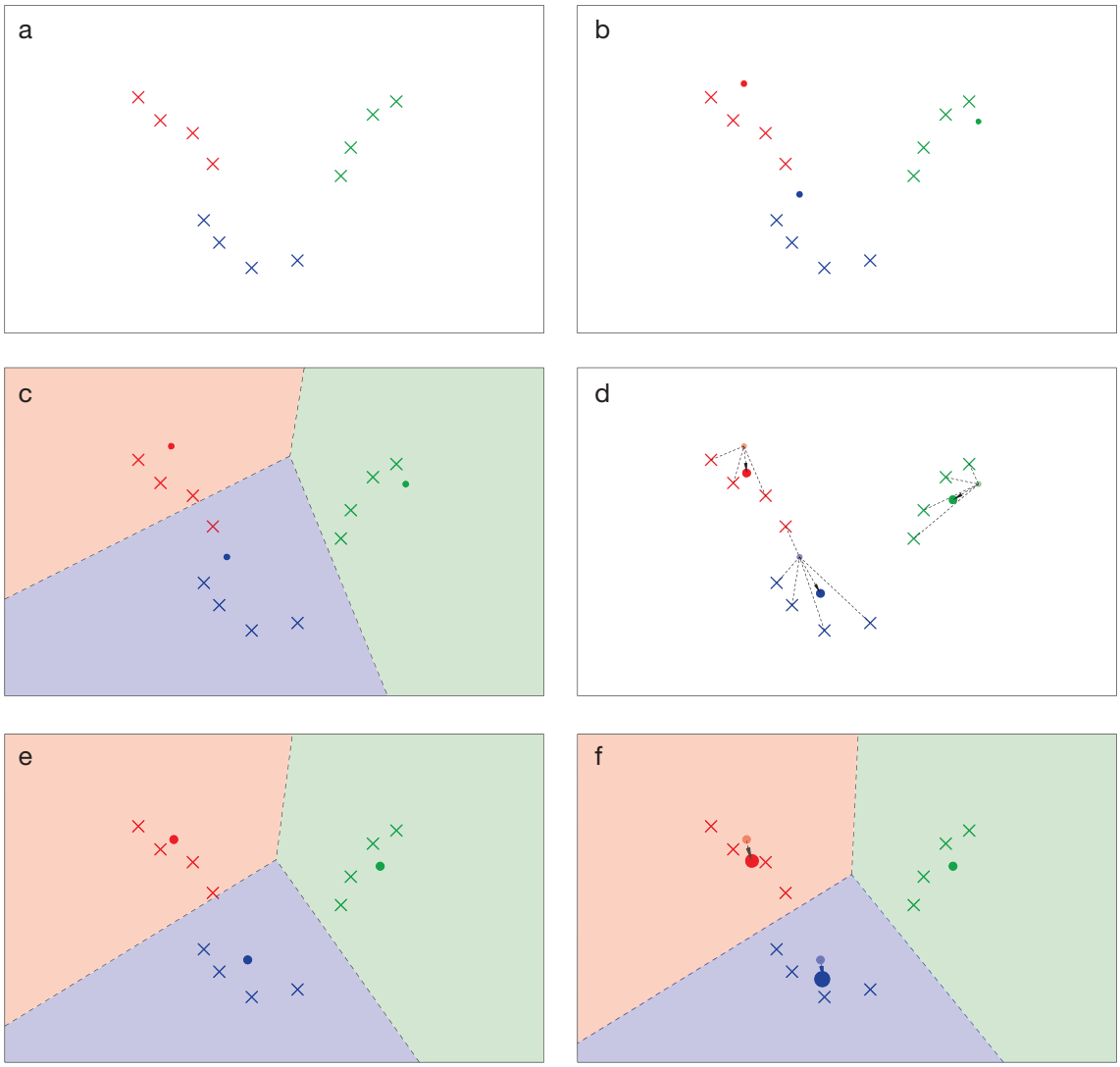


Figure 1.1 - *k*-means clustering.

The following 6 panels illustrate the process of *k*-means clustering. The above example displays a collection of 12 data points distributed in a 2 dimensional feature space, with a single feature along the horizontal and vertical axis. There are 3 groups within the dataset represented by a different colour. The starting locations of the centroids can either be explicitly specified or randomly determined. Since *k*=3, there are three centroids, represented as a filled circle, in this scenario (b). The algorithm then proceeds to assign each data point to the closest of the three centroids using a series of perpendicular bisectors (c). After all data points have been assigned, the centroids' locations are then shifted to the mean location of the corresponding centroid groups (d and e), and the process is repeated until no further displacement occurs (f).

Consider a dataset distributed in 2-dimensional space with one feature along the *x* axis and another along the *y* as illustrated in figure 1.1. The value of *k* refers to the number of starting points (centroids) the algorithm will use to explore the feature space. This in essence specifies the maximum number of clusters you expect to find within the dataset. The example in figure 1.1 uses a *k* value of 3. The starting locations of the centroids can either be specified by the operator or randomly determined. Next the data points are then assigned to their closest centroid by using a series of perpendicular bisectors resulting in the formation of 3 clusters based on the entire dataset. For each cluster the associated centroid is updated to the mean of its constituent data points. The process is then repeated, iteratively moving each centroid to the minimum distortion point (MacKay, 2003) with termination of the process defined by a minimum displacement. In this way the algorithm does not necessarily need to evaluate all $\frac{n(n-1)}{2}$ pair-wise dissimilarities, where *n* is the number of data points.

To evaluate the displacement between iterations, a method to calculate the distance is necessary

$$d(p,q) = \frac{1}{2}\sum_{i}(p_i - q_i)^2$$

Variables used:

$$x^{(n)} = data\ point$$

$$m^{(k)} = set\ of\ centroids\ (or\ means)$$

$$k^{(n)} = current\ centroid\ (or\ mean) of\ cluster$$

$$\hat{k}^{(n)} = centroid\ (or\ mean) the\ data\ point\ is\ believed\ to\ belong\ to$$

$$r_k^{(n)} = indicator\ variable\ (identifies\ which\ centroid\ a\ datapoint\ is\ assigned)$$

Each data point is assigned to the nearest centroid within the set of centroids using the distance measure above.

$$\hat{k}^{(n)} = arg\min_{k}\{d(m^{(k)}, x^{(n)})\}$$

$$r_k^{(n)} = \begin{cases} 1 \;\; if \;\; \hat{k}^{(n)} = k \\ 0 \;\; if \;\; \hat{k}^{(n)} \neq k \end{cases}$$

After assigning each data point to its closest centroid, the means are adjusted to match the sample means of the data points they are responsible for, i.e. the locations of the set of centroids are updated.

$$m^{(k)} = \frac{\sum_n r_k^{(n)} x^{(n)}}{R^{(k)}}$$

Here $R^{(k)}$ is the total responsibility of the mean *k*.

$$R^{(k)} = \sum_n r_k^{(n)}$$

The process is then repeated until there is no further change in location for the set of centroids.

Although the algorithm tends towards a local minimum, it may not necessarily be a global minimum (Kanungo et al., 2002). In fact, depending on the starting point, the algorithm can discover a variety of solutions. To increase the probability of finding the optimal solution it is recommended to perform multiple runs of the algorithm at different start locations (Bradley and Fayyad, 1998; Duda and Hart, 1973).

These problems partly arise from the multi-dimensionality of the data and how the algorithm processes this information. Although it is possible to manually screen the images first (thereby forcing the algorithm to be at best semi-automated), this assumes the algorithm will always be able to correctly identify normal tissue and cluster these data points into one group. However *k*-means

clustering is known to have important limitations. These include its inability to represent the size, shape, weight or breadth of each cluster (MacKay, 2003). Therefore successfully clustering normal tissue into a single group is unlikely to be the prevailing result, since the component clusters within the dataset are heterogeneous in terms of these features.

Another drawback of the *k*-means method is hard-clustering, whereby each data point is assigned to exactly one cluster and all points within are equal in that cluster. Intuitively, it would appear more appropriate if data points located between 2 or more clusters played a partial role in determining the centroids of all the clusters it could plausibly be assigned to. To address this criticism the soft *k*-means algorithm was developed (MacKay, 2003).

Only the assignment step is modified to account for the "slackness" factor of the algorithm

$$r_k^{(n)} = \frac{\exp\left(-\beta d\left(m^{(k)}, x^{(n)}\right)\right)}{\sum_{k'} \exp\left(-\beta d\left(m^{(k')}, x^{(n)}\right)\right)}$$

This algorithm is similar to the original *k*-means formula, but possesses an additional parameter β. The parameter β represents how strict the algorithm handles its borders, such that as it approaches infinity, the more closely the soft *k*-means algorithm resembles the original *k*-means formula.

In spite of this, the necessity to specify *k* prior to execution remains a significant drawback. In a lesion segmentation application, *k* would be related to the number of lesions – one of the questions we are using lesion segmentation to answer. Removing the need to specify *k* would avoid the constraints and complications described above associated with *k*-means clustering. One alternative is the mean-shift clustering algorithm.

1.3.3.1.2 Mean-shift clustering

Mean shift clustering is a general non-parametric clustering procedure (Comaniciu and Meer, 2002). Unlike the *k*-means method, it neither requires prior knowledge of the number of centroids present in the dataset nor assumes a shape of the clusters. A search window (bandwidth), that isolates a specific volume within the *n*-dimensional feature space is positioned on the dataset. If each data point is given an equal unit weighting, the centre of mass (mean location) of the search window is calculated which will represent the location of the next centroid. The search window is then shifted so that the newly calculated centroid is at its centre (Fukunage and Narendra, 1975). The displacement vector is therefore dependent on the density gradient itself. This has two effects on the procedure. First, the vector will always point towards the direction of the maximum increase in density until convergence is achieved. Second, the algorithm automatically adjusts its convergence speed, with smaller steps as the window nears the maxima.

In this way, the dataset with points in an *n*-dimensional feature space is treated as a probability density function, where dense regions correspond to the local maxima (modes) of the underlying distribution. Each data point within the dataset is then processed using the algorithm. Those that share the same (at least approximately) maxima locations are considered members of the same cluster. The only parameter that is required before execution of the algorithm is the bandwidth. This is a significant benefit over *k*-means clustering, since if the bandwidth could be selected prior to execution it would facilitate a fully automated segmentation routine. However its selection is not a trivial matter. Larger bandwidths provide the opportunity for larger displacement vectors enabling the algorithm to identify maxima more rapidly, but at the expense of its resolution.

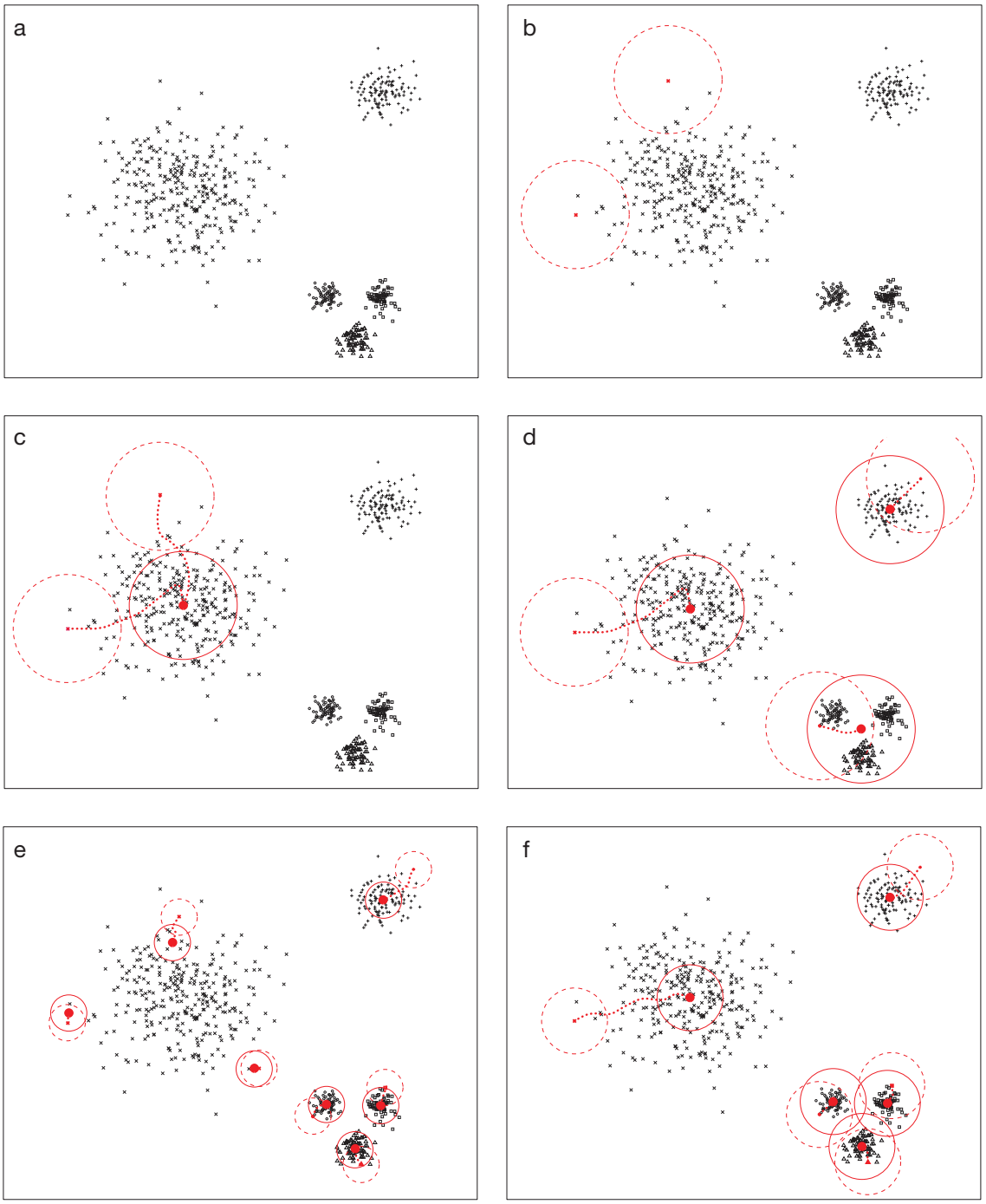


Figure 1.2 - Mean shift clustering.

A dataset of 5 clusters is distributed in a 2 dimensional feature space. Each datum is selected in turn as the starting centroid. A bandwidth is specified prior to running the algorithm and can be visualised as a circle, centred on the centroid, with all data points lying within its borders used to calculate the next centroid location (b). The starting bandwidth is represented by the dashed circle, with the course of the centroid depicted by the red dotted line with its final location (maxima) identified by the red dot. Data points who have the same final centroid are clustered together (c). Large bandwidths traverse the feature space swiftly but risk losing spatial resolution (d). Small bandwidths risk over-fitting (e). The optimal bandwidth is able to separate all the clusters (f).

In figure 1.2 we have a set of data points distributed in a 2 dimensional feature space. It is clear to see that there are 5 discrete clusters, with their constituent data points represented as a cross (+), x, circle, square or triangle, present in this space which we would like the algorithm to identify. If a large bandwidth is selected, the algorithm will quickly identify the 2 centroids (represented as a filled circle) for the cross and x clusters. However it will fail to separate the 3 smaller clusters in the bottom right hand corner because the bandwidth is too large, thus grouping these data points together with a common centroid. Conversely, if a small bandwidth is used, the algorithm will have greater focus on local features and ignore the gross structure, resulting in numerous inappropriate centroids being identified.

As the bandwidth is decreased, the maximum possible shift per iteration is also reduced. This will increase the computational load on the algorithm, as more iterations will be required before the algorithm identifies its maxima. Thus the challenge with mean-shift clustering lies in the preparation of the data to ensure the greatest contrast to assist classification with the largest bandwidth.

The situation is fairly simple if the various clusters (of interest – lesioned and unlesioned) are well separated and of a large magnitude. This is however a rare and unlikely case, since it is difficult to differentiate between lesion and signal artefact (for example) using signal intensity alone. Consequently there may be a significant number of data points that traverse or reside in saddle regions of the probability density function and thus will be particularly sensitive to bandwidth selection. One solution to this problem would be to select a range

of bandwidths, observe the various outcomes, and determine the optimal cluster arrangement using the various derived solutions.

Although referred to as a benefit, the ability of mean shift clustering to determine the number of clusters automatically presents another potential problem. Normal tissue is not homogenous, being coarsely divided into gray matter, white matter and cerebrospinal fluid (CSF). Consequently it is likely that normal tissue will not cluster into one single group, but actually comprise of a number of "splinter" groups. Post processing of the clustered image will therefore be needed, whereby lesioned and normal tissue clusters are differentiated. Indeed some algorithms (using a voxel based framework) use probabilistic map priors of the 3 aforementioned regions to determine whether a voxel belongs to any of these groups or to a separate "lesioned" group (Crinion et al., 2007).

In both described clustering methods, the mathematical calculations employed are not particularly complex. Despite this, the amount of data that needs to be processed in one brain volume forces both methods to still be time consuming. Moreover if a range of *k* values or bandwidths are to be investigated the amount of processing required per volume increases further.

1.3.3.2 Supervised learning algorithms

Supervised techniques are methods where there are already some predefined classes for a known *training* set of data. The algorithm *learns* the relation between the label and the data in the training set and is then used to determine the class of a new, *testing*, set of data. This prior knowledge can be in the form of instances of both classes, as is true for classification methods such as those based on Support Vector Machines and Relevance Vector Machines (Boser et al., 1992; Tipping, 2001) or instances of the "normal" class as with anomaly

measures (Cover and Hart, 1967; Harmeling et al., 2006; Van Leemput et al., 2001).

The process of gathering and incorporating training (or reference) set data can be difficult. Since this reference set is in essence the optimal classification result, any inconsistencies contained within it will be propagated by the algorithm with careful thought necessary to define the reference set to prevent the breakdown of the classification process. If not, the algorithm may identify an arrangement that has not been envisaged but still fulfils the specified criteria. This increases the difficulty in obtaining the data that fulfil the specific criteria sought, consequently reducing the size of the reference set and the algorithm's accuracy and reliability. Conversely, even where this issue is not a problem, large data sets place a significant computation burden on the method, as access to this detail is required in some form or another. One solution is to have a training period, where the algorithm "learns" the pattern from the reference set before applying it to the test data, consequently parameterising and summarising the pattern. In this way, most of the processor heavy calculations which are common amongst runs can be performed once, thus reducing computation time in the long term.

1.3.3.2.1 Support Vector Machines

Support Vector Machines (SVM) are a group of training algorithms that maximize the margin between sub groups of interest within the dataset (Cortes and Vapnik, 1995; Hsu et al., 2003; Vapnik, 1982). The algorithm attempts to generate a flat plane, commonly referred to as the *hyperplane*, which separates the two sub groups within an *n*-dimensional feature space and provides the largest distance between the plane and the nearest data point from each sub group. Once trained the resultant model can be applied to a novel test datum,

whose classification will depend on where it falls in relation to the hyperplane (Boser et al., 1992).

There is a trade-off between the number of adjustable parameters of a classifier and the generalizability of the solution. As the number of parameters increases, the classifier has a greater chance of learning the training set without error. However this comes at the cost of reducing the eventual generalizability of the solution to *de novo* data (Geman et al., 1992; Vapnik, 1982). SVM handles this problem by automatically tuning the capacity of the classifier by searching for the maximum margin. Although SVM is capable of handling high dimensional data, for simplicity, low dimensional examples will be used to provide an overview of its theory.

1.3.3.2.1.1 The kernel trick

Consider a dataset consisting of 2 sets of values, represented in figure 1.3 as dots, filled and unfilled. The objective of the classifier is to separate the 2 groups with a single line or plane with the largest margin possible. In one dimension the solution can be easily found if the dataset is linearly separable (figure 1.3a). The task becomes more complicated when a non-linear solution is required (figure 1.3b). To overcome this SVM transforms the data into a higher dimensional (feature) space, using a kernel. The objective of this step is to enable SVM to discover a linear solution in the new higher dimensional (feature) space (figure 1.3c). Since the instructions used to transform the dataset from native space to higher dimensional feature space are known (the kernel), the algorithm is able to apply the inverse instructions to the linear solution in feature space and thus derive the non-linear solution in native space (figure 1.3d).

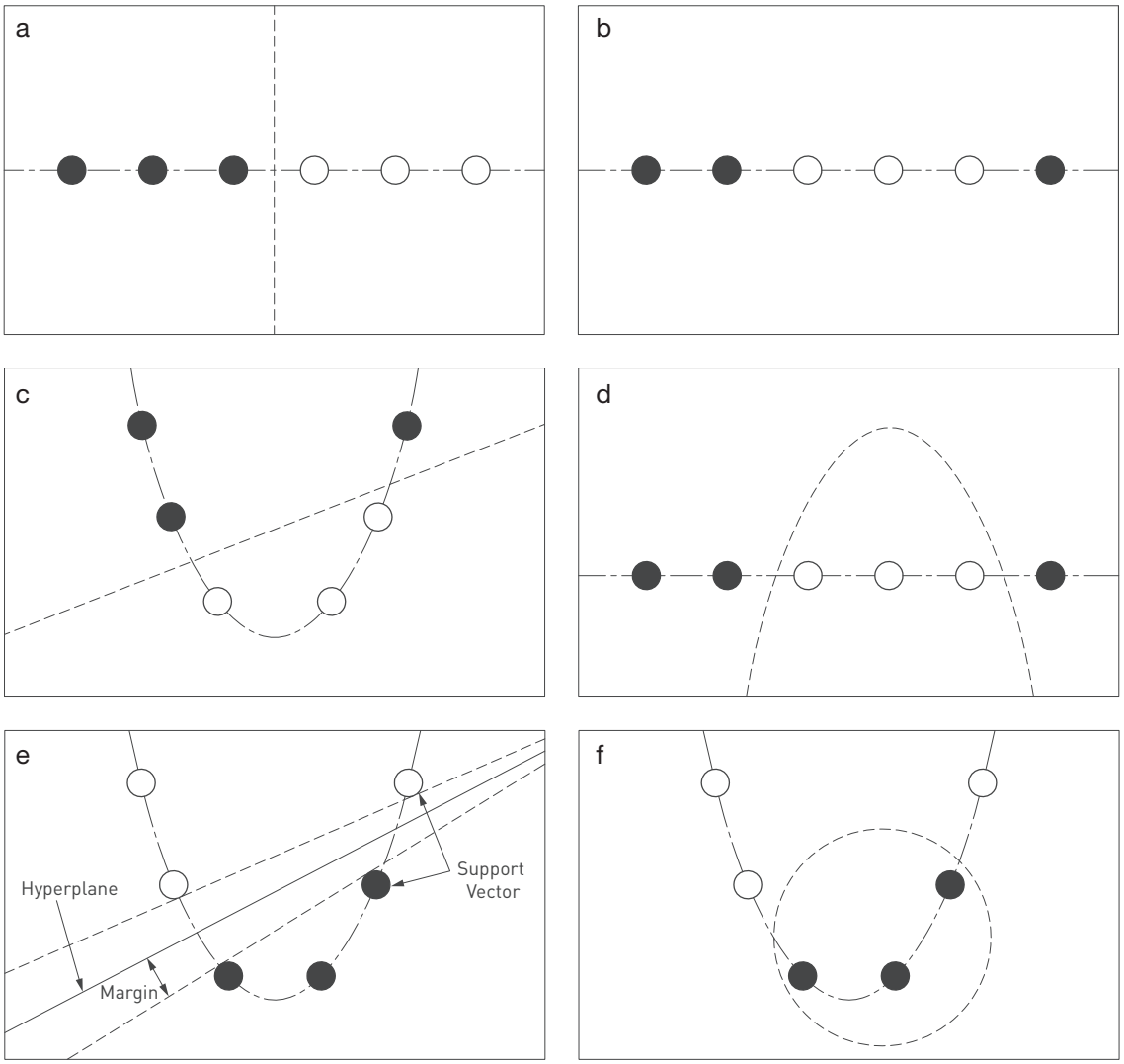


Figure 1.3 - Support Vector Machines (SVM).

The above example illustrates how SVM classifies a one dimensional data set. In a simple case, the data may be easily separable whilst remaining in one dimension (a). It is difficult to separate the 2 groups within the data set shown in (b) whilst remaining in one dimension. Unlike (a) a single linear line is not a feasible option, either 2 separate lines or some form of curve is required to separate these two groups. By using a kernel function, SVM transforms the data into a higher dimensional space, in this case 2 dimensions, to facilitate the identification of a linear solution (c). Since the instructions used to transform the dataset from native space into higher dimensional feature space are known (the kernel), the algorithm is able to apply the inverse instructions to the linear solution discovered in higher dimensional feature space and thus derive the non-linear solution in native space (d). SVM assesses the dataset and identifies a collection of data points that best describe the boundary between the two groups. These data points are known as support vectors, with the dividing plane called a hyperplane. SVM attempts to maximise the margin between the hyperplane and constituent support vectors (e). Alternatively rather than identify a dividing plane, the smallest sphere that best separates the data points can be found this is the basis of one-class SVM (f).

1.3.3.2.1.2 Maximum margin / Hyperplane

Although the dataset has been mapped into feature space there may be more than one solution to the classification problem.

To ensure only one answer, SVM defines the optimal solution as the maximum margin separation of the groups. Theoretically the optimal dividing line (hyperplane) is located between the two groups where the distance from the nearest data point in each group to the hyperplane is maximized.

Since the algorithm is trying to create a dividing plane between the two sub groups, it only needs to define the border between the two. Consequently the classification function depends only on those training examples (data points) that are closest to this boundary and where the hyperplane will eventually be positioned (figure 1.3e). These data points are referred to as Support Vectors (SVs).

The finer mathematical details regarding the SVM algorithm are not necessary for our purpose; however a general understanding of it process will help explain how it may be used for the task of lesion segmentation. A more comprehensive description of the mathematics behind the SVM algorithm can be found in (Boser et al., 1992).

Consider the example in figure 1.3. The data have been mapped into feature space where the hyperplane is located at:

$$w \cdot \varphi(x) + b = 0$$

The vector that transforms the data points into hyperspace is denoted by φ, with *w* and *b* representing adjustable parameters. The distance between the outer margin and the hyperplane is *M* such that the margin width is:

$$margin\ width = \frac{2|M|}{\| w \|}$$

Consequently the optimal solution for the hyperplane is where the margin width is maximised.

$$maximum\ margin = max \frac{2|M|}{\| w \|}$$

$$subject\ to\ (w \cdot \varphi(x) + b \geq M \quad \forall x\ of\ class\ 1$$

$$subject\ to\ (w \cdot \varphi(x) + b \leq -M \quad \forall x\ of\ class\ 2$$

In this 2 class SVM example, we can set the upper margin and lower margin to one and minus one respectively, such that:

$$class\ 1: \ w \cdot \varphi(x) + b \geq 1$$

$$class\ 2: \ w \cdot \varphi(x) + b \leq -1$$

$$maximum\ margin = max \frac{2}{\| w \|}$$

$$maximum\ margin = min \frac{1}{2} \| w \|^2$$

$$subject\ to\ (w \cdot \varphi(x)) + b \geq M \quad \forall x$$

The parameters *w* and *b* for the optimal solution can be found where there is a unique minimiser to the quadratic equation.

When a linear kernel is used, it is possible to extract the vector *w*. Importantly, each element, or weight, in the vector is associated with a specific feature or dimension. The polarity of the weight indicates which class the dimension is important for. The magnitude of the weight describes the importance of the dimension in relation to the other dimensions in the model, with larger absolute values dictating a greater influence to the classification process.

Above describes the 2 class situation where the transformed data is linearly separable. However, even after transformation the data may still be non-linearly

separable. To tackle this problem, akin to hard and soft clustering, SVM can be adapted to use soft margins.

1.3.3.2.1.3 Soft margins

It is unlikely for a kernel to completely separate all datasets presented to it, either owing to noise or to overlap of the distributions. Consequently, if strict criteria were set (hard margins) then it is conceivable that a single datum would render the algorithm incapable of cleanly separating the two groups, thus causing the algorithm to stop and fail. Instead a slack variable (ζ) can be introduced to allow some instances to fall within the margin but penalizes them if they do. In this way the algorithm is able to continue with its calculations and search for the most appropriate hyperplane that separates the two groups with the least number of errors.

A Support Vector Machine model, based on a linear kernel, requires the *C* parameter to be specified prior to training. The *C* parameter of the Support Vector Machine is a regularisation term that represents the trade off between the number of classification errors incurred on the training data and margin maximization (Rychetsky, 2001). Increasing the *C* value increases the penalty incurred by a misclassification error. Thus as *C* tends towards infinity the soft margin algorithm tends towards a hard margin algorithm (Rychetsky, 2001).

If the 2 groups within the data set are easily separable in the high dimensional feature space – such that there is a strip of feature space that is not inhabited by any data points – as the *C* value is increased, the accuracy of the model should remain constant as there is no change in the number of errors when the slack variable (ζ) is reduced. Conversely, if the 2 groups are poorly separable

the errors of classification will change as the *C* value is varied which will then have a greater chance of affecting the final prediction accuracy of the model.

Alternatively, the *C* parameter can be viewed as a “tuning knob” that balances the the trade off between the complexity of the decision rule and the frequency of errors (Cortes and Vapnik, 1995). The *C* parmeter therefore adjusts how tightly or loosely the data is fitted.

1.3.3.2.1.4 Kernel selection

Kernel selection is one of the most important parameters to setting up an SVM. The kernel implicitly defines the structure of the high dimensional feature space, where the maximum margin hyperplane is to be found. Too rich a feature space and there is a risk of over-fitting the data, and conversely, a hyperplane may not be found if the feature space is too poor (Cristianini et al., 1998). Many kernels have been proposed by researchers in the literature, the four most commonly discussed are, linear, polynomial, radial basis function (RBF), and sigmoid. The radial basis function (RBF) kernel is recommended as the first-line kernel (Hsu et al., 2003), as it performs best when averaged across various data sets, though it is acknowledged that with data sets where the number of features / attributes is very large, a linear kernel may perform better and more efficiently. The RBF kernel non-linearly maps samples into higher dimensional space, so unlike the linear kernel can handle the case when the relation between class labels and attributes is non-linear. Also under certain conditions and parameters, both the linear kernel and sigmoid kernel will behave like a RBF kernel (Keerthi and Lin, 2003; Lin and Lin, 2003). The RBF kernel also only requires the *C* and gamma parameters to be adjusted, whilst

the polynomial kernel has more hyperparameters that need definition prior to accurately classifying the data.

Applying SVM to lesion segmentation, a 2-class model may be used, where class one would consist of voxels representing normal tissue and class two, lesioned. Once the classifier has been trained, SVM is a computationally efficient method, able rapidly to process and segment a *de novo* volume. Much of the difficulty with the use of SVMs originates from training the classifier. The performance of the derived solution is dependent on the available training data, which will need to be imaging modality specific. The manual identification and collection of a large normal dataset is relatively easy. However, for the lesioned dataset it is not quite as simple.

First, each brain image in the set must be manually identified along with segmentation of the inherent lesion. Ideally lesion coverage by the set should not only involve the entire brain, but with sufficient representation at each voxel in its lesioned form: inevitably a very labour intensive process. Second, the training data supplied to train the classifier will influence how the resultant model can be later applied. As mentioned earlier, the classifier will have been tuned to work most effectively for the same imaging modality as the training data. This is a somewhat minor issue; however, extrapolating this argument, it is likely that the classifier will be most efficient for the type of lesion used in the lesioned training data. Take for example magnetic resonance diffusion weighted images (MR dwi) of stroke patients. Generally speaking fresh lesions are visualized as bright white regions. However on occasions these lesions are not homogenous, consisting of a bright white ring with a black haemorrhagic core. It is possible that the classifier will have difficulty with voxels representing these haemorrhagic cores as it has been trained primarily on data that describes lesions as bright white. Third, a reference dataset implies co-registration of the images to allow comparisons across the whole brain. Spatial

normalisation techniques have been originally developed to handle healthy, unlesioned brains. The abnormal signal of lesions interfere with the process of spatial normalisation thereby deteriorating the registration process and increasing the noise introduced into the system.

Up to now, two class SVM methods have been discussed where the two classes, lesioned and normal tissue, have been explicitly defined in the training stages. Alternatively, we can simply define what is normal, and anything that deviates from this definition is considered abnormal and consequently lesioned tissue. This is the basis of a one class SVM method, where only the classification of normal tissue is required. Support vector data description is an extension of the one-class SVM algorithm, where instead of a hyperplane describing the boundaries between groups, a hypersphere describes the multi-dimensional volume that encapsulates the features of the class (Tax and Duin, 1999).

The geometry of the sphere is described with a radius whose size is the minimum required to contain all (or most of) the data points about a known centre. Data points that fall outside the sphere are considered outliers (figure1.3f). The algorithm is designed to derive the optimal values for the sphere's radius and centre, with the tightness of fit guided by kernel selection and adjustment of spread / width of the kernel.

Extending the concept of using outliers to determine whether a voxel is lesioned or normal tissue, one can consider simply indexing a voxel's likelihood of being an outlier.

1.3.3.2.2 Anomaly measures

Outliers are observations that are numerically distant from the rest of the data; however there is no rigid mathematical definition for this, with their

identification ultimately being a subjective exercise. Nevertheless it is generally accepted that an outlier would represent an entity that was detached from the main distribution. Consequently, it would appear appropriate to consider this concept for the use of lesion segmentation, where lesioned tissue is an excursion from the usual distribution. Often with normally distributed data, outliers are identified as data points that lie a specific number of standard deviations from the mean. Colliot et al, applied this concept to the detection of focal cortical dysplasia on T1 MR images, where voxels possessing a signal intensity more than 4.5 standard deviations from the mean were identified as outliers (Colliot et al., 2006). Alternatively, rather than comparing the actual value of the data point to a reference value of the normal dataset, e.g. the mean, outlier determination can be based on some derived surrogate marker of distance from the normal dataset. The Mahalanobis distance (Mahalanobis, 1937) was adapted for segmentation of multiple sclerosis lesion on MR images by van Leemput (Van Leemput et al., 2001). A Mahalanobis distance was calculated for each test voxel with respect to its stereotactic identical voxel dataset, in the reference normal brains, identifying those above a specified threshold as outliers.

Extending the use of distance metrics, a dataset may be interrogated by the search for distance-based outliers. Using the theory of nearest neighbours, methods such as *k* nearest neighbour provide an ordering (index) of the data points according to their typicality, with atypical points being labelled as outliers. Moreover these nearest neighbour methods can index a dataset that enable a non-parametric method for outlier detection in a comparatively fast manner (Anbeek et al., 2004, 2005; Harmeling et al., 2006).

1.3.3.2.2.1 *k*-nearest neighbour (*k-NN*) density estimator

The *k*-nearest neighbour density estimator, assess the density at a particular point *x* by deriving the smallest sphere centred on *x* which contains its *k*-nearest neighbours (Cover and Hart, 1967; Fix and Hodges, 1951; Harmeling et al., 2006). Consider a 2 dimensional (2D) feature space, with a different feature represented by the *x* and *y* axis. Within this feature space resides our reference data represented by unfilled dots. A test data point can be placed in this feature space and its likelihood of being an outlier determined by its proximity to the reference dataset. The *k*-nearest neighbour density estimator calculates the radius of a sphere that represents the distance between the test point and its k^{th} nearest neighbour. This metric is usually the Euclidean distance between two data points, though other metrics can be used.

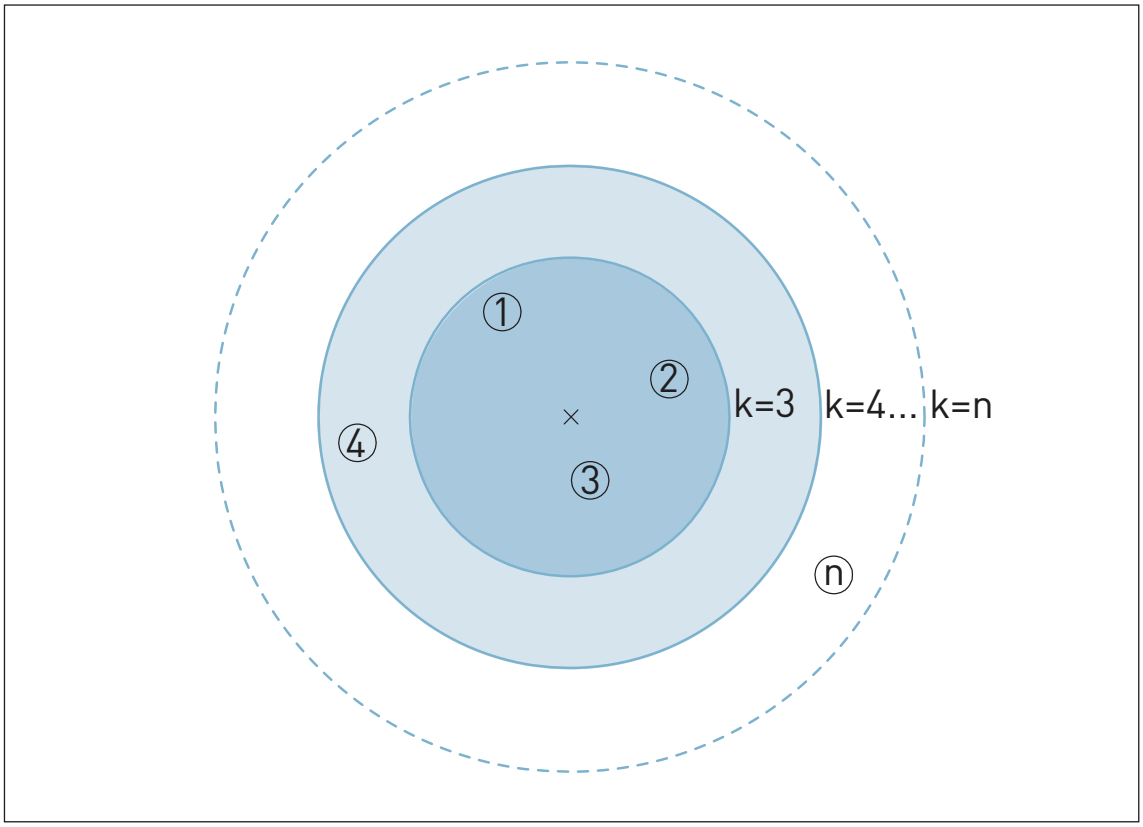


Figure 1.4 - *k*-nearest neighbour (*k-NN*).

The data point *x* is surrounded by 4 satellite data points in a 2 dimensional feature space. The distance of the furthest satellite point from *x* within the "clique" of *k* nearest neighbours determines the value of the *k-NN* score.

The rationale underpinning the *k-NN* is the assumption that outliers reside in sparse regions. Thus the magnitude of the index value is inversely related to the density of the local region around the test data point.

The mathematical simplicity of the nearest-neighbour rule provides a fast, non-parametric method that can easily handle high dimensional datasets. Also, for a fixed *k*, the error rate of this method approaches twice the Bayes error as the dataset tends towards infinity; whilst for a fixed dataset size, the error rate tends towards Bayes error rate as *k* approaches infinity (Cover and Hart, 1967). Despite its favourable error characteristics, increasing the reference dataset places greater computational demand on the method. These measures are forms of instance based learning, where the algorithm does not parameterise the reference data into model. Instead the reference data must be stored such that the algorithm may access the "raw" detail to identify the nearest neighbours each time it is run, thus deteriorating its speed benefit. Furthermore, the rate of reduction in bias is particularly slow with datasets of high dimensionality (Fukunaga and Hummels, 1987). Part of this error can be managed by increasing *k*, but caution must be taken as this may reduce the algorithm's ability to clearly identify the inherent classes (Cover and Hart, 1967). Limitations of *k-NN* is attributed to its apparent wastefulness as it only considers the largest distance within its *k* nearest neighbours, and ignores the remaining distances to the closer neighbours within the sphere (Harmeling et al., 2006).

1.3.3.2.2.2 Gamma (γ) density score

The gamma score refines the *k-NN* method by averaging all the distances of the *k* nearest neighbours. In other words gamma is the mean distance to its *k* nearest neighbours.

$$\gamma_k(x) = \frac{1}{k}\sum_{i=1}^{k} d(x, nn_i(x))$$

When the reference dataset consists of several clusters with different variances, the value of *k*, in both the gamma and *k-NN* methods, determines the level of focus on local properties. The adjustment in the algorithm for the γ score reduces its vulnerability to anomalous data points included in the group of *k* nearest neighbours.

1.3.3.2.2.3 Zeta (ζ) anomaly score

Gamma is however still density dependent and is susceptible to problems if there are significant density differences between clusters within the reference dataset – dense regions will have low values and sparse regions high ones – making the labelling of anomalies density dependent (Harmeling et al., 2006).

This is illustrated in figure 1.5a, where a heterogeneous synthetic two-dimensional dataset is labelled by the gamma score of each point: the larger the diameter of a circle the higher the gamma score. It is easy to see that members of the smaller, denser cluster will generally have a smaller value of gamma than members of the larger, sparser cluster even though they belong equally strongly to their respective clusters. Thus, if anomaly were to be determined by a fixed threshold of gamma, either too many members of the sparse cluster will be labelled as anomalous or too few of the denser one: it is easy to see that the score is confounded by local differences in density.

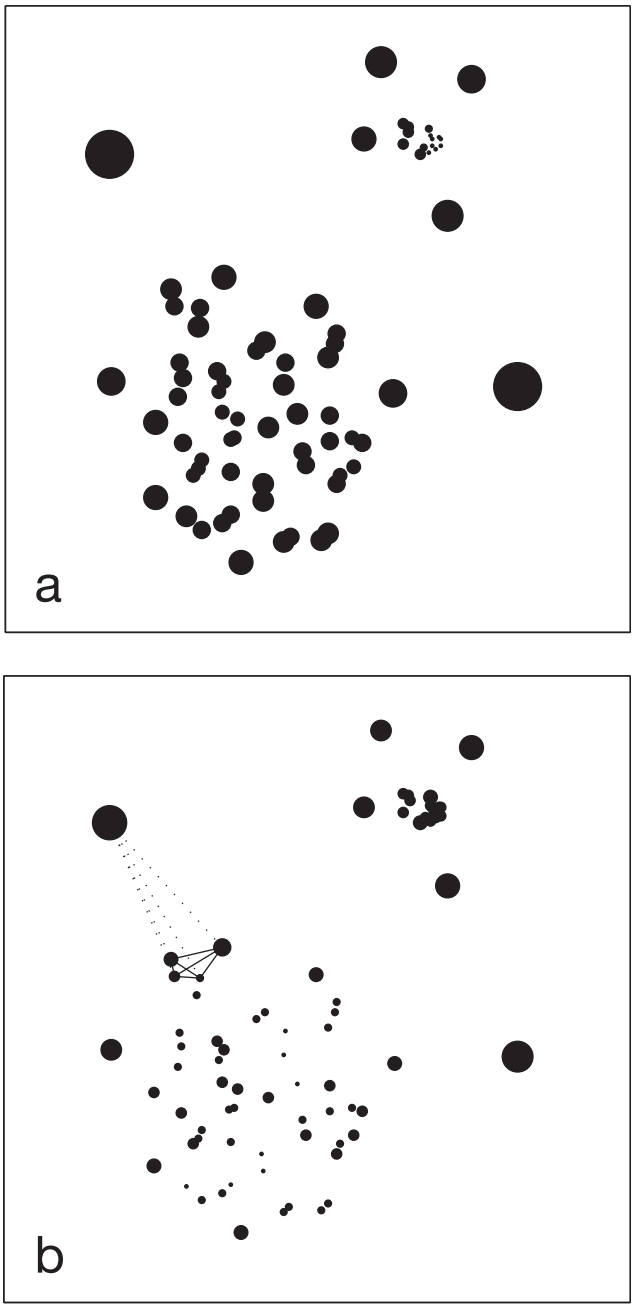


Figure 1.5 - Gamma (γ) and Zeta (ζ) anomaly scores.

The above diagram displays a collection of data points dispersed in a 2 dimensional feature space. There are 2 main clusters, with a dense cluster in the top right corner surrounded by 4 satellite outliers, and a less dense cluster in the bottom left corner. The diameter of each point is proportional to its anomaly score. The points have been indexed with the gamma score in (a) and the zeta score in (b). If a threshold was set to separate the upper right cluster from its 4 satellite data points, with the gamma score many of the data points in the bottom left cluster would be identified as anomalous rather than belonging to a group of its own. This problem ameliorated by indexing by the zeta anomaly score.

If one considers the “clique” of *k*-nearest neighbours and its inner relations, however, this problem can be reduced. The idea, metaphorically put, is that no matter how strong your friendships are you are an outsider if your best friends are closer to each other than to you. The zeta (ζ) score (Rieck and Laskov, 2007) is thus the difference between the average distance from a point to its neighbours (an index of the anomaly of the point in relation to its neighbours) and the average inner-clique distance of its neighbours (an index of the density of the neighbourhood clique):

$$\zeta_\kappa(x) = \frac{1}{k}\sum_{i=1}^{k} d\big(x, nn_i(x)\big) - \frac{1}{k(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{k} d(nn_i(x), nn_j(x))$$

As described above, the first part of the formula (on the left of the subtraction sign) is essentially the γ score for the *k* nearest neighbours. The second half of the formula (on the right of the subtraction sign) represents the mean distance of the clique of *k* nearest neighbours.

Figure 1.5b shows the same synthetic data labelled by ζ, with the calculation for one point graphically illustrated: the first term is indicated by the dotted lines, the second by the solid lines. Comparing the two clusters, it is easy to see that in contrast with γ, variations in density do not disturb the score here.

As discussed above the application of a density measure on the task of lesion segmentation is not a new idea. It provides a non-parametric technique, which is mathematically simple, and requires no prior learning on training data. Also by using a voxel based method, location is coded by the voxels’ stereotactic position and not in the data presented to the algorithm, thus keeping dimensionality down. Data storage and memory still remains an issue, coupled with greater computation time, with large reference datasets being required to minimize error. However the adaptation of the zeta anomaly score to this task may help to offset some of the computational constraints whilst improving on this class of lesion segmentation algorithms that use the *k-NN* rule. Also the

relative ease, with which the Zeta anomaly score can be adapted to handling more dimensions (features), opens up a novel route of lesion segmentation where information from more than one image modality can be incorporated and processed, by the same algorithm, to classify the data at the voxel level. These features, along with the need to specify relatively few parameters prior to its execution, lend the zeta anomaly score favourably to a fully automated segmentation algorithm. As with all the techniques discussed that create a map of the brain indexing a voxel's probability of being a lesion, the goal is to emulate the gold standard of manual segmentation. Therefore a binary image inevitably needs to be generated to allow objective comparisons, with some surrogate feature guiding the selection of this threshold. The use of the zeta anomaly score for lesion segmentation is introduced and evaluated in §2.

1.4 Spatial normalisation

The concept of comparison assumes prior knowledge of identity so that one knows what one is comparing. To make spatial comparisons between matching anatomical labels, we must first ensure that homologous regions of the brain are in register with each other through a process referred to as spatial normalisation (Ashburner and Friston, 1997, 1999; Brett et al., 2001; Friston et al., 1995). Normalisation is achieved by transforming the brains into a common stereotactic space that may either be a group mean image or a standard template such as Talairach space or the Montreal Neurological Institute (MNI) template (Ashburner and Friston, 1999; Brett et al., 2001).

Spatial normalisation cannot be perfect because the correspondence between individual brains is not perfect. Spatial normalization sets out to minimize the amount of anatomical variability between individuals, compensating for differences across individuals. At the gross level, the human brain exhibits significant consistency in its overall structure across individuals despite considerable variability in its size and shape. Each human brain normally consists of two multilobed hemispheres grossly symmetrical around the mid-sagittal plane, exhibits a set of fairly consistent major sulcal landmarks, and contains a very consistent set of deep brain structures. However on closer inspection the finer details of the brain structure, eg. gyral size and number, displays greater variability (Thompson et al., 1996). The objective of spatial normalisation is therefore to appropriately minimize these differences.

1.4.1 Automated spatial normalisation algorithms

Manual labelling of brain regions is an established approach for establishing anatomical correspondence, however the expenditure in time and resources often limits its application (Klein et al., 2009). To facilitate studies, even with

modest sample sizes, automated techniques have been developed. There are four popular approaches to spatial normalization: landmark; surface; volume and computational anatomy. All four techniques rely on the consistent gross structural features across all brains to help guide the transformations that are designed to reduce the anatomical variability of the finer details.

1.4.1.1 Landmark based methods

Landmark based spatial normalization methods rely on gross physical features of the brain to determine how the brain should be warped into a standard space (Poldrack, 2011). This approach was employed by Talairach where anatomical landmarks including the anterior and posterior commisures, mid-sagittal plane and exterior boundaries of the brain in each direction were used. Although tools and programs that employ landmark based methods still exist they have generally been superseded in popularity by volume or computational anatomy techniques.

1.4.1.2 Surface based methods

The cerebral cortex can be considered to be a simple 2 dimensional sheet where sulci and gyri are essentially folds within the sheet. Surface based methods extract various morphological metrics such as sulcal depth and cortical convexity across the entire brain surface to calculate the transform necessary to warp the brain to a standard template (Fischl et al., 1999; Tosun et al., 2004). Many of these methods are semi or fully automated and use a far greater number of parameters to derive the transformation function compared with traditional landmark approaches. However, it is acknowledged that there is considerable variation between secondary and tertiary sulci across individuals, despite the more stable similarities of primary sulci and their associated cortices. For example the sulcus of Jensen, double cingulate sulcus (also

known as the paracingulate sulcus) and supra-orbital sulcus are not present in all individuals, therefore violating the direct one-to-one correspondence and complicating the registration task (Pantazis et al., 2010). As the name suggests surface based methods do not utilize information located in deep brain structures. Consequently this technique has often found popularity with those interested in functional imaging where projecting areas of activation on a two dimensional plane more informative and visually intuitive, such as for example in the visual cortex.

1.4.1.3 Volume based methods

Volume based methods, sometimes referred to as non-label based approaches, identify a spatial transform that optimizes some voxel-based measure between the test and template image, where both are treated as an unlabelled continuous process (Ashburner and Friston, 1999). The derived transforms can either be linear – affine – where the transformation is applied to the entire brain or non-linear where a basis function is applied to a sub-region of the brain to improve the correspondence between the test and template image.

The templates utilised by these techniques can either be a standard template such as the MNI brain template, or bespoke to better suit the task. The spatial transformation is therefore guided by contrasts such as surface boundaries between brain and air, as well as deep sub cortical structures. In both instances, the algorithm aims to minimize the difference between the test and template image, as such this technique is susceptible to poor starting estimates (Frackowiak, 2004).

1.4.1.4 Differentiable homeomorphism

Differentiable homeomorphism is a special kind of transformation that can be represented as a 3 dimensional vector field (Poldrack, 2011). Each voxel location has its own unique vector that maps its displacement from the original to the transformed image. Consequently these methods have a huge number of parameters, but are regularized to ensure that the deformations are smooth and do not violate the topology of the structures being transformed (Ashburner, 2007). Owing to the increased number of parameters that are being modelled these methods are computationally expensive. Although they take significantly longer to run, they are still significantly faster than manually transforming the image. However recent studies comparing the performance of diffeomorphic techniques with volume based methods have shown superior performance of the former (Klein et al., 2009).

1.4.2 Spatial normalisation of lesioned brains

The aforementioned registration techniques inevitably rely on prior knowledge either in the form of another brain or a template. Almost invariably the latter will consist of one or more brain volumes that is representative for its target group, devoid of any focal abnormality. Although a bespoke template can try to capture some of the morphological changes associated with age related atrophy, the same is far more difficult for focal lesions.

The pattern of focal lesions is extremely heterogenous with respect to morphology, distribution and signal intensity making the creation of a satisfactory template of this kind near impossible. Therefore in its absence, if we attempt to register a test image against the MNI template, intuitively, the

registration procedure will perform at its optimum in the absence of a lesion as it has no reference data to guide where to place the abnormal tissue.

There are 3 general ways in which the effect of the lesion tissue may be minimised. The first involves exploiting the inherent symmetry present across the mid-sagittal plane. The brain volume is reflected in the mid-sagittal plane, and then a fusion image is created by calculating the mean of the original and reflected brain volume (Weiller et al., 1995). Although this approach dilutes the effect of the abnormal signal in the ipsi-lesional hemisphere, it potentially extends the effect of the abnormality to the contra-lesional side. The second and most widely used method is to completely ignore the abnormal signal when calculating the transformation parameters, in a process called normalisation with cost function masking (Brett et al., 2001).

The third approach, enantiomorphic normalization, uses the symmetry in the mid-sagittal plane to help correct the abnormality in the lesioned brain (Nachev et al., 2008). The ideal situation would be to have a version of the brain volume immediately prior to the injury, however in most cases this is not a viable option. Due to the inherent symmetry of the brain, the contra-lesional hemisphere is our best record of what the ipsi-lesional hemisphere would have been prior to the injury. Consequently, rather than ignoring the abnormal signal and leaving a “hole” in the data, the signal intensities of the abnormal voxels are replaced by their corresponding (enantiomorphic) contra-lesional ones thereby providing more information to the normalization routine, while avoiding the propagation of the abnormality to the contra-lesional hemisphere. (Nachev et al., 2008).

1.5 Inference

The process of determining the relationship between the anatomy of the brain and a specific symptom or function is the last step in a lesion-symptom mapping study. The aim is to create a mathematical model of this relationship, from which we can infer how the brain is organised. The necessity for choosing the most appropriate model is not to accurately describe the collected data – this would be over-fitting – but to gain an insight into what the patient population is like, i.e. its generalizability, and to make predictions on future events. As we are trying to relate the architecture of the brain to a symptom, the model is therefore attempting to describe the relationship between the spatial information – the location of an injury – and the symptom under investigation. By examining the model and its parameters, a window into this relationship may be found.

Careful consideration must be taken over the statistical technique applied to the training data as it will impose various assumptions about the patient population from where the data originated. Unfortunately, we often do not know all of the parameters of the patient population, with only the information present in the collected data to guide us. In most situations it is possible to discern some of the parameters, and make an educated guess about how to model the data.

Broadly speaking there are two types of analysis; univariate analysis, where the effect of a single variable on an outcome of interest is modelled; and multivariate, where multiple dimensions are studied and the effects of all variables on the outcome of interest is assessed together. It has been shown that the human brain is a highly organised organ with an extensive inter-connected network of neurons. This would suggest a distributed system is at least possible (Berman et al., 2006; Bullmore and Sporns, 2009; Fair et

al., 2009; Greicius and Kimmel, 2012; Haxby et al., 2001; Hutchison et al., 2012). Despite this, the vast majority of past lesion symptom mapping studies have applied some form of univariate analysis to generate their inferences. In the following sections I shall briefly describe some of the popular univariate techniques used and then introduce a couple of multivariate methods which might be considered as an alternative.

1.5.1 Univariate methods

Past lesion studies have assessed a collection of brains and generally used a univariate approach to their analysis. Back in the 19th century Paul Broca conducted careful observations of patients and subsequent post mortem examinations on the brain to correlate the ventro-posterior region of the frontal lobe to speech production (Broca, 1861). Broca's conclusions were based on the visual assessment on a number of brains, noting the size and location of injury and documenting regions of commonality. At the heart of the deductive reasoning was the assumption that each unit location within the brain was spatially independent from its neighbours, i.e. the state of the unit location, healthy or damaged, did not influence the state of another unit location. This approach is essentially a univariate technique since the effect of a single variable (a unit location) on the outcome of interest is examined in isolation to the other variables.

The advent of computed tomography (CT) and Magnetic Resonance Imaging (MRI) removed the moral requirement to await the timely passing of the patient as it allowed the brain to be visualized, quantified and monitored *in vivo.* This not only increased the number of patients available to study, but helped increase the spatial resolution of analysis, improve temporal resolution (minimise the delay between the onset of injury and the measurement of its structural features) and provide a means to collect suitable control subjects.

Despite this the techniques used to perform lesion symptom mapping have not changed significantly and still employ a univariate approach that uses the same assumptions made by Broca.

Three of the most prevailing techniques are the template overlay, subtraction overlay method and voxel-wise lesion symptom mapping (VLSM) (Bates et al., 2003; Damasio and Damasio, 1989; Rorden and Karnath, 2004).

1.5.1.1 Template overlay method

When a brain is damaged by some focal pathological process, the pattern of injury bears no resemblance to the underlying functional architecture. Importantly, the scale of brain lesions is around 10^{-2}m, whilst in contrast the putative scale of functional organisation suggested by the spatial heterogenicity of individual neurons is in the order of 10^{-3}m. To make population inferences of the functional role of different parts of the brain, previous studies have often relied on comparing different sets of patients with large overlapping lesions.

The template overlay method collects a cohort of brain images with a paired outcome of interest. A binary mask of the lesion within the brain image is created and all masks are then brought into register with each other, to ensure homologous regions of the brain are aligned. Subsequently, these masks of relatively large lesions are superimposed one on top of another to identify critical loci much smaller than each individual (Karnath et al., 2001; Rorden and Karnath, 2004). Similar to a Venn diagram the overlapping lesions create a region of peak overlap that is common to all. It is thus inferred that damage

to this common region is critical to the development of the impairment shared across the group of patients.

Differential vulnerability is a specific sub-type of problem that can arise without controlling for confounds and is a recognised complication that can produce erroneous loci. This is where certain brain regions are more prone to injury than others i.e. the locations of lesions are not randomly distributed within the brain, but are subject to various factors such as vascular supply and surrounding skull. Consequently the technique may identify a region that is not correlated to behaviour but instead a region related to the confound. This is of particular importance for lesion symptom mapping since the most commonly used lesion data is from patients with vascular brain injury. To help illustrate this problem, consider a moderately sized region *A* critical for a function *F*, where impairment is observed if 15% or more of region *A* is damaged. Next consider a neighbouring region *B* which is always injured as well when any part of region *A* is affected due to the property of differential vulnerability. Usually the region of maximal overlay is selected from the generated frequency overlay map, so despite both region *A* and *B* being injured, the level of overlap at *B* will be greater than *A*, thereby falsely localising the behaviour to lesions in region *B*. This example is illustrated in figure 1.6.

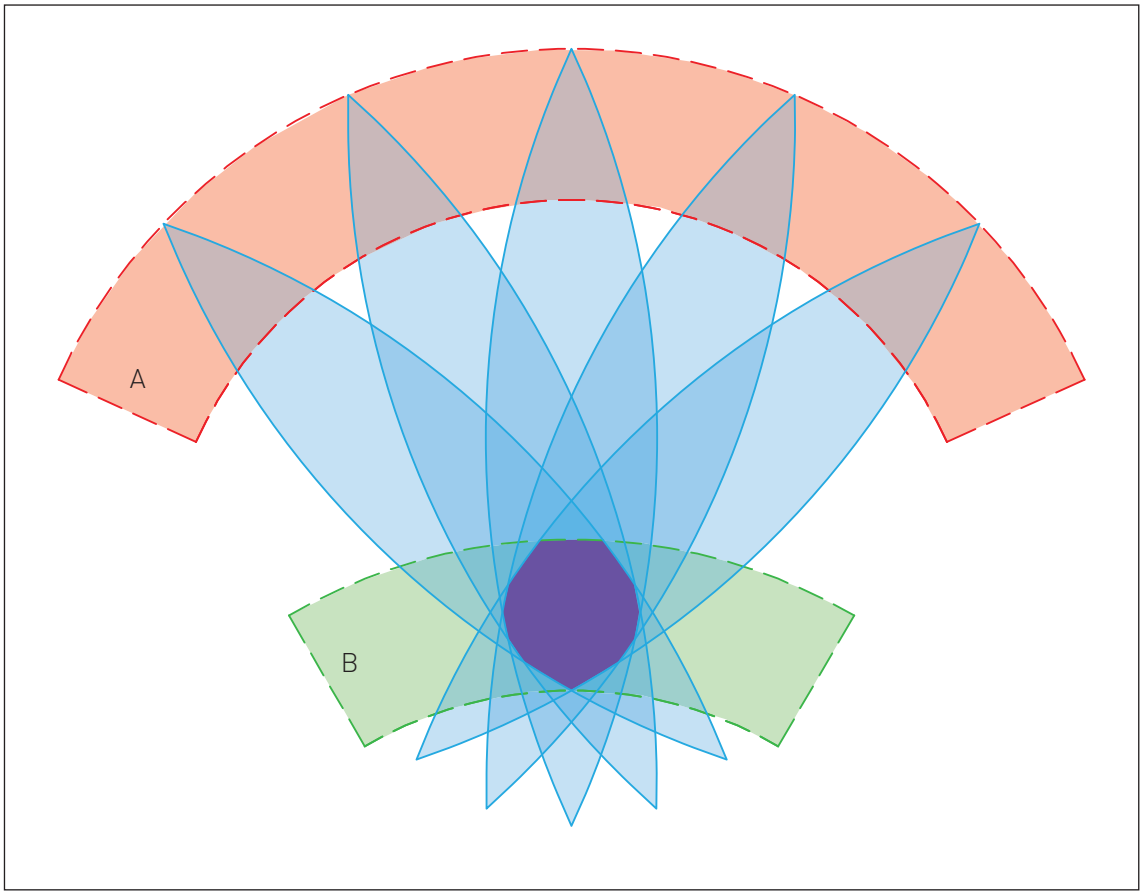


Figure 1.6 - Template overlay method.

Illustration of how stereotyped patterns of brain damage (schematized in blue) across a set of patients can hypothetically mislocalise damage of any part of critical area *A* (red) to the non-critical area *B* (green). This will happen whenever the spatial variability of damage to a non-critical area is less for the group or factor of interest than for the critical area. Such stereotypy of damage – a hidden deep structure in the data – may occur where the lesions follow a consistent non-neural architecture, as is the case with vascular lesions.

Owing to the spatial variability of injury secondary to some confound being less than the correlation between our critical region *A* and feature of interest, the overlapping lesions, represented by the blue petals, results in a higher level of overlap in region *B* rather than critical region *A*.

1.5.1.2 Subtraction overlay method

The subtraction method can be considered as an extension of the template overlay method, with the modification aimed at reducing the effect of confounds. Two groups of patients are selected: the first being a group of individuals who all have the same specific disorder and the second a matched control group (Karnath et al., 2001; Rorden and Karnath, 2004). The selection filter for the control group is defined as patients who are identical to the first group except for the absence of the specific disorder. A difficult task particularly when they must possess a lesion but without the same or new deficit! The overlay map of the matched controls is *subtracted* from the overlay map of the patient group, with both maps being created using the template overlay technique described earlier. The control group then acts to be a source of noise and bias secondary to the technique, which is then subtracted out from the analysis, therefore hopefully distilling the correlation of interest. This technique appears reasonable for noise, however it relies on the biases within the data to be well behaved, i.e. to remain consistent despite changes in the collection criteria, an assumption that cannot be guaranteed.

1.5.1.3 Voxel-based Lesion Symptom Mapping (VLSM)

An alternative approach has been proposed by Bates et al, (Bates et al., 2003) where the analysis is performed in a voxel-wise manner. Much like the subtraction method there are 2 groups of patients identical to each other except for the presence or absence of a specific impairment. Each voxel is

attributed 2 binary labels: healthy or injured; and with or without impairment, thus creating 4 distinct classes. In the paper by Bates (2003), the authors perform a *t*-test at each voxel, so that unlike the overlay and subtraction method, this technique interrogates each voxel independently to evaluate the level of contribution its injury will have on the functional outcome.

This mass univariate approach will then generate a brain map whose voxels are labelled with a *t*-score. Tackling the problem in this manner allows a finer resolution of analysis, where the characteristics of the continuous behavioural score may be captured and lesion boundaries are limited to voxel resolution. In addition it is possible to pass the data to an ANCOVA to remove the effect of a known confound. Bates et al used this ability to counter the possibility of a diaschetic effect – where a region may emerge secondary to highly correlated lesions some distance away – and minimize the probability of mislocalisation. They illustrated this technique with two parings: the insula with Broca's area; and Wernicke's area with the middle temporal gyrus. In both cases it was postulated that one of the areas in the pair only became significant secondary to a diaschetic effect caused by its corresponding partner. In the former pairing, voxels in the insula remained significant after factoring out Broca's area, whilst the reverse was not the case when the insula was factored out. Similarly in the latter pairing, after factoring out Wernicke's area the voxels in the medial temporal gyrus remained significant, whilst the opposite was true when the medial temporal gyrus area was covaried out. This technique does provide significant advantages over the template and subtraction method. First of all there is an inherent control group with VLSM. In addition, it is possible to minimise the interference from *known* confounds with the use of an ANCOVA, assuming the data permits its isolation.

As alluded to earlier, the selected model places certain constraints on the data. If we take a simple scenario such as a coin toss, this may be modelled

using either a normal, Poisson or binomial distribution as long as our sample size is sufficiently large enough to render the skew of the parent population minor. However, common to all three choices is the necessity for the data in the set to be independent – the result of one event is not correlated to the result of any other event. If this criterion is violated, none of the models proposed would be appropriate. Returning to lesion symptom mapping and the VLSM technique, a *t*-test is chosen to model the relationship at each voxel location. As a consequence it assumes that the state of one voxel – whether healthy or injured – is independent from the next. The model therefore only looks at the relationship between the presence of a symptom and the state of the voxel in question at each execution. Of course this is repeated hundreds to thousands of times due to the total number of voxels within the brain to create the brain map; however at its core it is still only performing a huge number of simple tests which uses a single variable – mass univariate analysis.

The benefit of its simplicity is the speed of computation and the tractability of analysis. Passing a few hundred of binary masks through this analysis will only take a couple of minutes to process on a standard workstation computer, with the output result providing a probabilistic map of the brain areas that are correlated to the symptom. Indeed many of the past lesion studies have employed such a method to accommodate the limitations of the available computing power (Bates et al., 2003; Dronkers et al., 2004; Geva et al., 2011; Verdon et al., 2009). By using a mass univariate technique such as the *t*-test or Fisher-exact test, we have immediately assumed that each voxel is spatially independent of one another, with any violation of this assumption nullifying the validity of the statistics. Unfortunately we cannot guarantee this independence, particularly with vascular lesions where the site and shape of the lesion is

heavily dependent on the pattern of the underlying vascular tree (Dronkers et al., 2007).

Ideally we would like to model the data, where each voxel within the brain is not assessed in isolation to one another, so the presence or absence of a symptom in a patient is modelled by the state of *multiple* voxels at the same time – multivariate analysis. This approach, as would be suspected, is computationally more intensive, and increases exponentially as the number of variables – voxels – increases.

1.5.2 Multivariate approaches

Intuitively it can be seen that our models are stronger if they are able to integrate more of the available information. In this way we can try to minimize the effect latent biases and account for the potential interactions between the different dimensions. This hopefully will improve the robustness of our inferences but it should be noted that it may have no bearing on the predictive power of the model. Ultimately the predictive power will be dependent on the true underlying relationship between the input data and the outcome we wish to model.

Understandably, many of the studies in the past exploring lesion symptom mapping have shied away from multivariate statistical analysis due to the computational cost of such calculations, however there has been an increasing adoption of multivariate approaches as these costs shrink with the rise of cheap computing power (Friston et al., 2008; Smith et al., 2013). Below is only a selection of 3 techniques that can be adapted to the problem of lesion symptom mapping and is not an exhaustive list.

1.5.2.1 Support Vector Machines

Recently Support Vector Machines (SVM) (Vapnik, 1982) have been gaining popularity in the field of neuroscience, especially as a classification tool (Mourao-Miranda et al., 2011; Plant et al., 2010). The concept of the technique is to use kernel based algorithms to transform the data into a higher dimensional space to simplify the derivation of a classification hyperplane. Although it is possible to extract the data points – support vectors – that are used by the algorithm to derive the classification function, the weight of each voxel towards the classification process is only possible with a linear kernel.

Although the ability to quantify the contribution of each voxel is not essential to the performance of the classification model, this detail may provide a window into which features / dimensions are more important for the task of classification.

For example, consider each voxel in a brain volume as a separate dimension. Next using a set of lesioned brain volumes and associated outcome data such as the ability to walk we can attempt to differentiate between those individuals who can walk and those who cannot, based on their lesioned brain image. The weight of each voxel's contribution to the classification model describes how important that location is to determining the presence or absence of the function in relation to the other dimensions, thus providing an insight into which areas are believed to be critical. It should be noted that the support vectors identified define the classification function by modelling the boundary between the different classes, rather than the areas typical of each class. Further information regarding Support Vector Machines can be found in §1.3.3.2.1.

1.5.2.2 Relevance Vector Machines

Relevance Vector Machines (RVM) are similar to SVM, but instead uses a sparse Bayesian approach which enables it to offer a probabilistic solution to the problem. I will first provide a brief introduction to Relevance Vector Machines, before comparing it with Support Vector Machines. This introduction is not meant to be an exhaustive insight into RVM, but hopefully convey the concept of RVM to assist the understanding of its application to our situation. A more comprehensive description of RVM can be found in Tipping (2001)

Relevance Vector Machines are based on a Bayesian formulation of a linear model with an appropriate prior that result in a sparse representation. It is this sparseness that facilitates its speed of computation, by ignoring those dimensions whose relevances are deemed insignificant. The linear model can then be represented as:

$$y = wx + c$$

Where *w* is the parameter vector, *c* is the offset and *x* is the input values used to predict the outcome *y*. Generally the offset *c* is incorporated into the vector *w*. If the relationship between *x* and *y* is non-linear then a kernel function can be used.

$$y = w\emptyset(x)$$

In this case $x \mapsto \emptyset(x)$ is a non-linear mapping — a basis function.

In our arrangement, we are trying to derive *w* – the weights – from our training data. The assumption here is that our training data is representative of our true model y_i, albeit with some additional noise. Thus our function can now be written as:

$$\begin{aligned} t_i &= y_i + \varepsilon_i \\ &= w\phi(x_i) + \varepsilon_i \end{aligned}$$

The assumption for ε_i is assumed to be a set of independent samples from a Gaussian noise process with zero mean and variance σ^2, such that

$$\varepsilon_i \sim N(0, \sigma^2) \ \forall_i$$

Therefore the probability of an outcome t_i given an input x_i with our model should be:

$$P(t_i|x_i, w, \sigma^2) \sim N(y_i, \sigma^2)$$

$$P(t_i|x_i, w, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(t_i - w\phi(x_i))^2\right\}$$

Ideally we wish to incorporate all our training data. To do so we can represent each training data point, t_i – the outcome values – in a vector t, with an associated design matrix Φ such that the last row in the matrix represents the vector $\phi(x_i)$. The design matrix merely contains the different basis functions, $\phi(x_i)$, at all the training points, t_i, for each of the weights in the vector w.

$$P(t|x_i, w, \sigma^2) = \prod_{i=1}^{N} (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2\sigma^2}(t_i - w\phi(x_i))^2\right\}$$

$$P(t|x_i, w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} exp\left\{-\frac{1}{2\sigma^2} \parallel t - \Phi w \parallel^2\right\}$$

Where:

$$t = t_1 \cdots t_N$$

$$w = w_1 \cdots w_M$$

There are M weights associated with the algorithm at initialization. As alluded to earlier, smoother functions, and thus less complex functions, are generally more resilient to over-fitting and result in better generalization. By applying constraints on the number of weights, we are in essence applying a smoothing term, thereby reducing the risk of over-fitting. This is achieved in the form of a prior on the weights, with a zero-mean Gaussian distribution.

$$P(w|\alpha_i) \sim N(0, \alpha_i{}^{-1})$$

Here α_i describes the inverse variance — the precision — of each w_i. Therefore there is a separate α_i associated with each weight, modifying the strength of the prior.

To make predictions using the Bayesian model the posterior probability, over all the unknown parameters, given the data needs to be computed. This probability cannot be computed analytically because of its complexity, and approximations need to be made. First decompose the posterior probability to:

$$P(w, \alpha, \sigma^2|t) = P(w|t, \alpha, \sigma^2) \cdot P(\alpha, \sigma^2|t)$$

Rearranging and substituting β^{-1} for σ^2

$$P(w|t, \alpha, \beta) = \frac{P(t|w, \beta) \cdot P(w|\alpha)}{P(t|\alpha, \beta)} \sim N(\mu, \Sigma)$$

Where μ is the mean and Σ is the covariance of the probability distribution given by:

$$\mu = \beta \Sigma \Phi^{\mathrm{T}} t$$

$$\Sigma = (A + \beta \Phi^{\mathrm{T}} \Phi)^{-1}$$

$$A = diag(\alpha)$$

In order to derive μ and Σ we need to find the hyperparameters α and β which maximises $P(\alpha, \sigma^2|t)$.

Since $P(\alpha, \sigma^2|t) \propto P(t|\alpha, \sigma^2)P(\alpha)P(\sigma^2)$ we can ignore $P(\alpha)$ and $P(\sigma^2)$ if we assume they are uniform hyperpriors. This simplifies the problem of maximising the evidence to:

$$P(t|\alpha, \sigma^2) \equiv P(t|\alpha, \beta) = \int P(t|w, \beta)P(w|\alpha)dw$$

At this point we need to substitute back our variables and then simplify. The following few steps are beyond the scope of this thesis and are not necessary for understanding of this algorithms application. Further details can be found in Tipping (2004). We eventually end up with:

$$P(t|\alpha, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \left(\frac{1}{2\pi}\right)^{\frac{M}{2}} \prod_{i=1}^{M} \alpha_i^{\frac{1}{2}} \exp\{-E(t)\}(2\pi)^{\frac{M}{2}} |\Sigma|^{\frac{1}{2}}$$

Where

$$E(t) = \frac{1}{2}(\beta t^T t - \mu^T \Sigma^{-1} \mu)$$

The above equation is the marginal likelihood, and by taking the logarithm becomes the log marginal likelihood and it is this equation which we need to maximise with respects to α and β, through a process known as the evidence approximation procedure.

$$\ln P(t|\alpha,\beta) = \frac{N}{2}\ln\beta - E(t) - \frac{1}{2}\ln|\Sigma| - \frac{N}{2}\ln(2\pi) + \frac{1}{2}\sum_{i=1}^{M}\ln\alpha_i$$

If we differentiate the above equation by α and β, we can generate 2 equations that isolate α and β in an iterative expression.

$$\alpha_i = \frac{1-\alpha_i\Sigma_{\mathrm{ii}}}{\mu_i^2}$$

$$\beta_i = \frac{N-\sum_i 1-\alpha_i\Sigma_{\mathrm{ii}}}{\| t-\Phi\mu \|^2}$$

In this way the α and β which maximise the marginal likelihood are found iteratively, starting with an initial value for α and β to calculate a *m* and Σ, and using the the latter two to derive new α and β. Also, α represented a constraint on our weights, *w*. As a consequence of the evidence approximation, many α values will tend towards infinity. Consequently, the variance, Σ, and mean, μ, of the posterior distribution for the corresponding weights will tend towards (0,0), and so the corresponding basis functions $\Phi(x_i)$ can be removed from the overall design matrix Φ at each iteration.

By setting a threshold criterion for the amount of change in α_i the iterative process can be terminated once no further significant change occurs.

Finally once convergence has been achieved — or the α_i criterion reached — we have:

$$P(t|x',\alpha,\beta) = \int P(t|w,\beta)P(w|\alpha,\beta)dw$$

$$P(t|x',\alpha,\beta) = N(\mu^T\phi(x'),\sigma^2(x'))$$

Where *x*' represents the new input datum. Therefore the estimate for *t* is the mean of the above distribution, $\mu^T \phi(x')$.

1.5.2.3 Comparisons between SVM and RVM

In both methods the algorithm tries to reduce the classification model down to a few vectors, with fewer vectors facilitating a faster calculation. With Relevance Vector Machines the derived weights are a reflection of the relative level of influence a dimension has on the classification model with respect to all the other dimensions. Those dimensions whose weights do not fulfil the criteria of significant influence are pruned by setting their weights to zero. The remaining non-zero weights are called the relevance vectors. In the 2 class classification situation, the magnitude of the relevance vector is proportional to the importance of that dimension to the process of classification, with the polarity directing it to one of either class. As such the vectors that have survived pruning are located *far* from the boundary margin.

In comparison, the support vectors identified with Support Vector Machines (SVM) also describe the classification model. However unlike RVM each support vector (SV) does not represent a different dimension, but instead represents a specific datum from the training set. In the 2 class classification problem, when the data has been transformed into a higher dimensional space the support vector helps define the location of the hyperplane, with all its associated dimensions. There will therefore be 2 groups of vectors – one for each class – which lie either side of the hyperplane and consequently define the border between each group. In addition each identified SV has an associated coefficient whose polarity dictates the class the SV belongs to, whilst the associated weights provide a guide to the relative importance of

the SV to the classification problem in the higher dimensional space. As a consequence, making inferences from the classification model is difficult.

As explained earlier, the optimal hyperplane is achieved by maximising the margins between the hyperplane and the support vectors. Similarly with relevance vector machines, the optimal solution is derived by iteratively maximizing the likelihood margin for the posterior distribution.

Although the generation of the classification function with RVM is slower than SVM, the parameters are automatically derived through the iterative method (although a threshold value will need to be specified for the minimum relevance magnitude), whilst a grid search for the optimal parameters is need with the SVM approach.

1.5.2.4 Bayesian Inference with Markov Chain Monte Carlo sampling

One of the main benefits of a sparse Bayesian method such as RVM is the comparative speed of computation compared with a full Bayesian approach. Conversely, although the weights associated with the relevance vectors provide an indication of the importance of each dimension – in this case our voxels – the algorithm is constructed to try and minimize the final number of dimensions resulting in the vast majority being pruned to zero. Although the RVM model has a relatively good level of prediction, the trade off for improved speed is a reduction in the localization of the critical regions.

Full Bayesian techniques maintain all dimensions allowing for a more detailed picture of relative importance between dimensions. As with most if not all Bayesian approaches is the difficulty in obtaining the posterior distribution. This generally requires the integration of high-dimensional functions which cannot

be solved analytically, with the alternative computational solutions presenting a considerably challenging hurdle.

In the following section I will focus solely on Markov Chain Monte Carlo (MCMC) methods. Once again this is not meant to be an exhaustive description and a more detailed explanation can be found in (Neal, 1993).

Markov Chain Monte Carlo (MCMC) methods are computational techniques that rely on random sampling to try and simulate direct draws from some complex distribution of interest. As the name suggests there are 2 main parts to the technique, Markov chains and Monte Carlo integration. Their appeal is that they ought theoretically to converge to the true solution in every case, and that they are relatively robust to over-fitting. These advantages come at great computational cost, as we shall see.

1.5.2.4.1 Monte Carlo integration

Originally the Monte Carlo approach was devised by physicists to use random number generation to compute integrals (MacKay, 2003). First consider a complex integral:

$$\int_a^b h(x)dx$$

This can be decomposed into the product of a function and a probability density function $p(x)$ defined over the interval (a,b).

$$\int_a^b h(x)dx = \int_a^b f(x)p(x)dx$$

As a result the integral can be expressed as an expectation of $f(x)$ over the density $p(x)$.

$$\int_a^b h(x)dx = \int_a^b f(x)p(x)dx = E_{p(x)}[f(x)]$$

The Monte Carlo simulation draws an independent and identically distributed (*iid*) set of samples from a target density distribution $p(x)$ defined by the set of possible configurations of a system. By drawing a large number of random variables $x_1 \cdots x_n$ from the density $p(x)$ then the equation can be approximated as:

$$\int_a^b h(x)dx = E_{p(x)}[f(x)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

This is referred to as Monte Carlo integration. This arrangement is particularly useful as it provides a means to derive the posterior distribution required for Bayesian analysis. For example to derive the posterior distribution, with a given prior the normalising factor in Bayes' theorem needs to be computed:

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

The normalisation factor integral can therefore be approximated by the Monte Carlo method as follows:

$$I(y) = \int f(y|x)p(x)dx = \frac{1}{n}\sum_{i=1}^{n} f(y|x_i)$$

Similarly the marginal posterior, $p(x|y) = \int_z p(x,z|y)dz$, can be solved using the same method. Therefore the crux of the method relies on drawing

(pseudo-) random samples from a specified probability distribution to estimate the intractable integral. To achieve this there are a number of prerequisites:

1. Probability distribution functions (pdf). This is the target distribution that must be specified by a set of pdfs
2. A random number generator
3. A sampling rule. This is a prescription for sampling from the specified pdfs.

The simulation can only apply the law of large numbers if the samples are independent (the average result will tend towards the expected result as the number of samples tends towards infinity). Satisfying this criterion may be difficult with Monte Carlo methods, however one solution is to use a Markov chain.

1.5.2.4.2 Markov Chains

A Markov chain is a mathematical system that undergoes transitions from one state to another in a chain-like manner (Neal, 1993). The key feature is that it is a random process, with the next state depending only on the current state. Although time is usually treated as a continuous variable, in this case time is considered to exist as discrete steps with the system occupying a specific state at each step and changing randomly between them. The changes in state are called transitions, each with an associated transition probability. The set of all states and transition probabilities completely characterizes the Markov chain.

Critical for its application to MCMC methods are three further features. First it needs to be aperiodic, such that the chain of transitions does not get trapped into a cycle. Second it needs to be irreducible, so that for any state of the Markov chain there is a positive probability of visiting all other states.

This ensures the transition matrix which defines the transition probabilities between states cannot be reduced to separate smaller matrices. This property is sometimes referred to as the transition graph being connected. Lastly, the Markov chain must have a stationary distribution. Thus irrespective of what initial distribution was used, the chain will eventually stabilize to this stationary (equilibrium) distribution.

1.5.2.4.3 Markov Chain Monte Carlo methods

Markov chain Monte Carlo methods therefore sample from a probability distribution based on the available data, by constructing a Markov chain that has the desired distribution as its stationary distribution. It combines the Monte Carlo method for sampling randomness and the Markov chain method for sampling independence with its stationary distribution. Although the Markov chain is constructed so that it will eventually converge towards its stationary distribution, the number of steps required in the chain can be excessive. Therefore it is important to design samplers that converge quickly, and that we do not begin using these samples until convergence has been achieved, otherwise we will not have been sampling from our desired distribution.

1.5.2.4.4 Auto correlation functions and burn in periods

With all MCMC methods, we need to ensure we are close to approaching, if not achieving the stationary distribution of the Markov chain. Unfortunately the number of steps required to reach this point is unknown, with “well-behaved” Markov chains only necessitating a few tens of steps, whilst others demanding tens of thousands. The steps involved in the approach towards the stationary distribution are ideally discarded, and these steps are generally referred to as the “burn in” period. Although predicting the burn in period analytically is not possible there are techniques available to help guide its selection. This is of

particular importance, not only to ensure the stationary distribution is reached, but also to identify chains that are slow to converge, or worse, do not show signs of convergence so the parameters defining the Markov chain may be adjusted.

With an MCMC random walk, if you sample after every step, the samples will usually be highly correlated: one step produces a point in parameter space that is very close to the original point. Consequently, the true error is under estimated as it does not decrease with the square root of the number of samples since they are not independent of each other. To further illustrate this problem, consider taking 1 measurement of some quantity. The precision of the measurement can be improved by a factor of 10 if a further 99 measurements are taken. However if the original measurement were simply to have been written down a further 99 times, the precision will not have changed at all since no further new information has been obtained.

In order to obtain better results, the MCMC simulation must wander around the state space for a while (the burn in period) so that its current point is no longer closely related to the point it was at before — move one small step away from a point and you are guaranteed to be close to that point, however after many random steps you may be anywhere, near or far.

Returning to autocorrelation functions, one way to identify whether a sufficient number of steps have been taken is to choose an observable feature and see how closely correlated it is with the value of that feature at a previous state. It is hoped the trend exhibited by the feature is a reflection of the state itself. In this way an idea of the number of steps required before the amount of correlation decreases to a suitable level can be found and consequently provide a guide to the length of the burn in period and ensure that each measurement is statistically independent of each other.

1.5.3 Predictive tools

The models generated by our inferential statistics have been used to probe the spatial relationship between anatomy and function. Evaluation of the model's fidelity to the desired correlation has, up to now, remained within the realm of theoretical examination, since unless the true spatial correlations are known, an objective assessment of the spatial inferences cannot be performed. It is only if the model can be definitively tested by another method that we can be sure of its fidelity to true life.

We do, however, have a surrogate marker in the power of the model to predict the label from the imaging alone. For example if the model was able to predict the outcome with 100% success it would be reasonable to argue that it has identified an important correlation between some input feature and the resultant outcome. Of course, until the model is tested on a novel test data set which it has not seen during training, we will not know if its success is secondary to accurately modelling the true population or simply over-fitting.

A process of cross-validation is used to minimise the chance of over-fitting, with variations in the techniques aimed at optimising the use of the available data. The process in essence rotates the total dataset through different divisions of training and testing sets, ensuring the algorithm is only presented the training data such that the derived model is naive to the corresponding testing set. Consequently the process is sometimes referred to as rotation estimation. In the following sections I shall provide an overview of some of the various cross-validation techniques.

1.5.3.1 Cross-validation

The goal of our trained classification function is to accurately model the true population / dataset. Theoretically this can only be achieved if we were to train

the algorithm on the complete true dataset; a request that is usually impossible to satisfy. Therefore we must appease ourselves with a sub sample, knowing that as our sample size increases, we will eventually approach the true dataset.

By accepting this compromise, we are then faced with two major questions – is our sample size sufficient to approximate the true population; and how well does our classification function model the true population? The two questions are closely inter-twined with each other, however more commonly we are faced with a fixed sample size (the largest harvestable and manageable), and consequently left with the problem of optimizing the model.

The performance of the model can be viewed as how well the model generalizes to the true population. By quantifying the model's performance we can then proceed to optimize the model by 'tweaking' the various parameters. Unfortunately the assessment of a model's generalizability is particularly difficult when the best reference – the true population – is unavailable. With only a sub sample of the true population, cross-validation is a technique employed to approximate the differences between the sample data and the true population – the generalization error – and facilitate optimization.

Cross-validation requires the sample to be divided into 2 groups: a training group and a test group, such that the model can be trained and assessed by two separate datasets. Explicitly defining these two data sets is particularly important. This is because the best model for the training set may not be the best model for the true population. If the test and training groups were not completely separate, our optimisation process would be exposed to the problem of over-fitting.

1.5.4 Assessing model generalizability

To illustrate this problem, figure 1.7 displays the same dataset classified by 3 different models. The model in figure 1.7a displays the property of under-fitting; figure 1.7c is an over-fitted model, while figure 1.7b is the optimal solution.

The classification function is depicted by the black line, with the data points classified according to which side of the line they fall. In the under-fitting scenario there is a substantial mix of red and blue dots either side of the classification line. Unless the training data is by chance an unfortunate unrepresentative sub sample of the true population the poor performance of the classification function on the training data is will most likely have a similarly poor performance on novel data from the true population.

At the other end of the spectrum is over-fitting. In this scenario the classification line has a far more complex course weaving between the dots as it passes through the training data. As a result the separation of the red and blue dots by this classification function is far more effective. Although this classification function has exquisite performance on the training data the complexity of the function probably reduces its generalizability to the true population, since the path of the classification line has been tightly constructed to fit the training data.

Although the basic procedure of cross-validation is as described above, there are a number of subtle variations. In all variations we have a training data set of size *N* consisting of the input and outcome data, further divided into a training group and an assessment group.

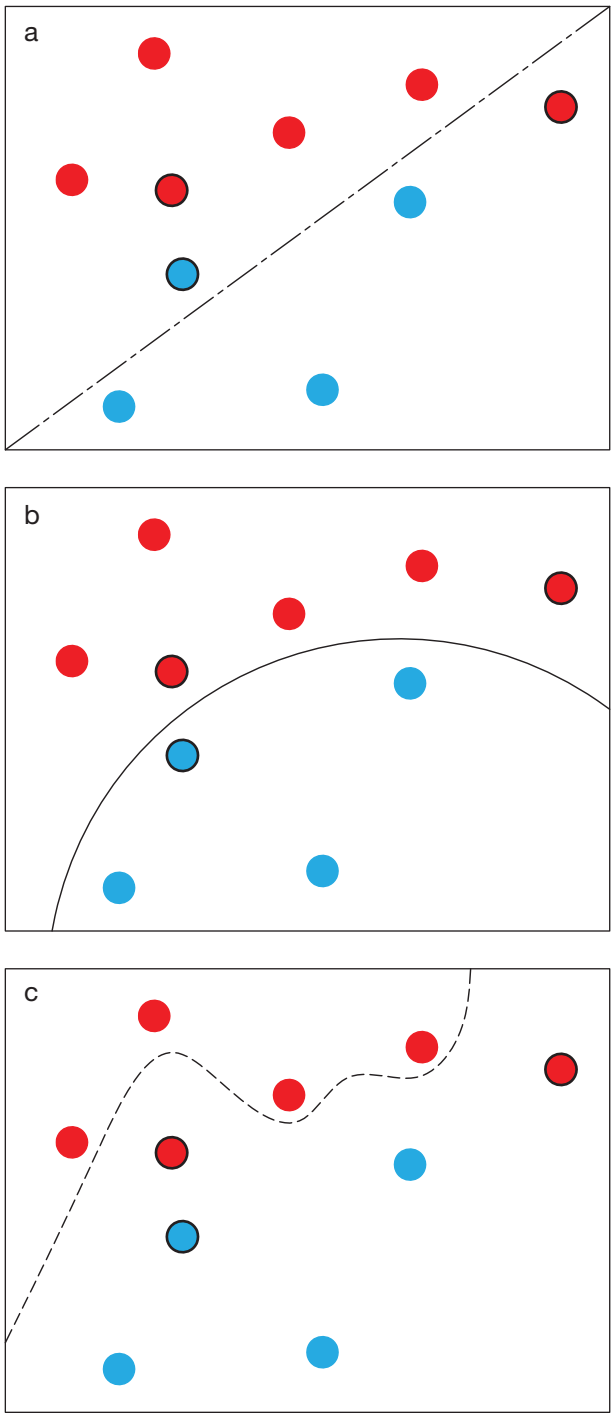


Figure 1.7 - Illustration of under-fitting and over-fitting.

Two clusters of dots distributed in a 2 dimensional feature space demonstrating under-fitting (a); optimal (b) and over-fitting (c). Dots with a black ring are test data, while those without represent training data. In (a) the classification function is very simple, however there are a number of misclassifications when assessed on the test data (under-fitting). In (c) the classification function has fitted itself very tightly to the red dots in the training data (over-fitting). When assessed on the test data the function appropriately classifies the blue dots but fails with the new red dots. A balance between these two extremes is the optimal solution with a relatively simple curve providing the best generalizability to the true population (b).

1.5.4.1 Leave-one-out cross-validation

In the leave-one-out variation, the assessment group consists of only one datum, whilst the rest of the training dataset is used to train the model, with the process repeated *N* times until each datum has been treated as an assessment datum (Kohavi, 1995).

This variation is particularly useful for continuous error functions, but conversely performs poorly with discontinuous error functions, such as the number of misclassified errors. In the latter case a small change in the data can cause a large change in the model. Consequently a *k*-fold cross-validation approach is preferred for tasks where we are assessing the number of correct or incorrect classifications by the generated model (see below) (Breiman, 1996).

This method should not be confused with jackknifing. In both cases, a single datum is excluded from the dataset, however with jackknifing the (*N*-1) dataset is used to estimate the bias of a statistic. In the jackknife some statistic of interest is calculated for each sub set of the data, with the average of these statistics being compared to the corresponding statistic derived from the entire dataset (*N*), thus providing a measure of bias.

1.5.4.2 *k*-fold cross-validation

The *k*-fold cross-validation method, divides the training dataset into *k* subsets, with the assessment group consisting of one subset (size: *N*/*k*), and the remaining data used for training (size: *N*-(*N*/*k*)) (Kohavi, 1995). Consequently if *k* were to equal the training dataset size, then the *k*-fold method would equate

to the leave-one-out method. Once again the process is repeated *k* times ensuring each subset has been used as an assessment group.

K-fold cross-validation is better suited to discontinuous data, and has the advantage over leave-one-out for model selection. It has been suggested that a *k* value equal to 10, where cross-validation is performed with 10% of the data, performs better than leave-one-out (*k*=1), while *k* values as small as 5 or 2 can work better if several different *k*-way splits of the data are used to reduce the variability of the cross-validation estimate (Breiman and Spector, 1992; Efron, 1983; Kohavi, 1995).

1.5.4.3 Split sample cross-validation

The crucial difference between split sample cross-validation and the previous 2 variants is in this case the data in the assessment group is kept constant and completely separate from the training process whilst in leave-*k*-out all *N* cases are used in training the model. Therefore the model is optimized using only the training group – being further divided into a sub training and sub assessment group – and the generalization error of the generated model is assessed using the assessment group. Therefore in split sample cross-validation the model derived from the data used for training is completely validated on a new sample, whilst with cross-validation the model is does not validate the complete training data, due to the mixing.

The drawbacks of this method are that it effectively reduces the size of the training data, as there is no crossover of data between the training and assessment groups. More importantly, unlike the leave-*k*-out variants, the split sample approach is in essence searching for the best division of the training dataset into the training group and assessment group not the lowest generalisation error.

1.5.4.4 Bootstrapping

Bootstrapping is a computationally intensive resampling method used to estimate statistics of the true population. Unlike traditional statistical methods the technique utilises the number-crunching power of computers to perform statistical analysis that lie beyond the reach of mathematical equations (Efron and Tibshirani, 1994). If we have a training dataset of size *N*, we will proceed to resample from this pool of data with replacement. Therefore if we decide to make each of our new resampled datasets of size *n*, we would randomly draw one datum from the training dataset, record its value, replace it back into the training dataset, such that when we randomly select the next datum we are still drawing from a pool of size *N*. This process is repeated until our resampled dataset is of size *n*. In this way we can generate numerous resampled datasets which help garner a more thorough analysis of the training dataset.

Although bootstrapping does not make any assumptions about the distribution giving rise to the data, there are a few pre-requisites that must be satisfied. First the available sample is a valid representative of the true population. This is critical since the assumption is that the population is to the sample as the sample is to the bootstrap samples. Second, the sub samples come from the same distribution of the population, but are drawn independently from the other samples (therefore the need for replacement).

Situations where bootstrapping is employed has been discussed by Adèr (Adèr, 2008). These include, when the theoretical distribution of a statistic is complicated or unknown, when the sample size is insufficient for straightforward statistical inference and when power calculations have to be performed and a small pilot sample is available.

With the above considerations, in this thesis I shall try to optimise models using the *k*-fold cross-validation technique. This is in part to reduce the

computational demand associated with the bootstrapping method and the discontinuous nature of the data.

1.6 Conclusion

The challenge of deciphering the functional anatomy of the human brain remains unanswered. Although a variety of new tools and techniques, such as functional MRI and transcranial magnetic stimulation, have been developed to probe the brain, limitations in their inferential power and spatial resolution compel a revival of historically the first method: lesion-mapping. In the past lesion studies have been plagued by modest sample sizes, but advances in computing give us the power to analyse large data sets with an ease almost comparable to fMRI studies. This thesis aims to present a potential solution to mining this data, from the processing steps through to the final inferential analysis.

In chapter §2 I shall tackle the problem of brain lesion segmentation. As mentioned earlier, the process of identifying lesioned from healthy tissue can be very time consuming. To help facilitate the preparation of a large data set, I propose a new unsupervised method for brain lesion segmentation assessed on a cohort of diffusion weight magnetic resonance imaging.

Chapter §3 addresses the issue of spatial normalisation of brain images. Here I shall assess the performance of a selection of spatial normalisation routines in the presence of a lesion.

Chapter §4 examines the potential for spatial mislocalisation when the data is parameterised and analysed in a univariate manner. Through a series of simulations I will show how univariate techniques maybe prone to mislocalisation and distort the lesion function inferences we draw.

Chapter §5 builds on the work described in §4 and compares the performance of a mass univariate technique (voxel-based lesion-symptom mapping) with a multivariate technique (Support Vector Machines) on a, more complex, multiple

loci model. I suggest that a multivariate approach may offer a solution to the mislocalisation issues surrounding univariate techniques and go on to assess three different multivariate methods on the same multiple loci model.

In the final chapter, I apply some of the methodologies to a real clinical situation and demonstrate how such work may not only explore the functional anatomy of the brain, but also act as a predictive model that can support clinical management.

2 A new method for unsupervised high-dimensional brain lesion segmentation*

2.1 Introduction

Functional neuroanatomical studies of the human brain have traditionally relied on relating a discrete area, or network of areas, to a specific function, or set of functions (Binder et al., 1997; Downing, 2001; Haxby et al., 2001; Shaywitz et al., 1994). The strongest evidence for such a relation is the observation of disruption of a function following disruption of its putative anatomical substrate. Although techniques such as transcranial magnetic stimulation (TMS) and deep brain stimulation (DBS) provide a means to achieve this they are spatially limited to accessible regions of the cortex. For a more comprehensive cover, we must rely on data derived from patients with focal lesions of natural, or incidental surgical, causes. Traditionally distinguishing damaged from healthy brain has been performed manually by eye; it is increasingly appreciated, however, that this approach is not only time-consuming (Andersen et al., 2010) but also potentially susceptible to operator bias (Ashton et al., 2003; Filippi et al., 1995). For these reasons a fully automated lesion segmentation algorithm that substantially outperforms manual methods in terms of speed and impartiality while remaining comparable in flexibility and accuracy is needed.

In general a lesion segmentation method must deal with 5 key problems. First, for any given imaging modality, the signal at any specific point in the

* A version of this chapter has been published in Cortex (doi 10.1016/j.cortex.2012.12.008)

brain will vary from one normal individual to another in a way that is difficult to parameterise. This variability is often not Gaussian, but multi-modal. It is therefore preferable to avoid making any assumptions of the distribution of the normal signal intensities and adopt a non-parametric approach (Lao et al., 2008).

Second, deciding whether or not a region is abnormal often depends on the signal not just in one imaging sequence but several different ones. Where the abnormality is not replicated across more than one type of sequence it may merely reflect noise or artefact. For example acute stroke lesions are generally represented as an area of high signal intensity on a b1000 sequence with a region of low signal intensity in the corresponding location of an apparent diffusion coefficient image (Warach et al., 1995). The potential to utilise multiple sequences in parallel is a significant benefit and would emulate the process of manual interpretation of the image more closely.

Third, although the signal properties of normal tissue may be definable, they are often not for lesions, simply because it is in the very nature of pathology to be heterogeneous in signal. Satisfactory priors therefore cannot be easily constructed without making unwarranted assumptions about the data that limit the generalizability of the algorithms based on them. Furthermore, even if one could capture the lesion signal distribution reasonably well, the main determinant of whether a voxel is labelled as lesioned or healthy is not its relation to the normal state but to the boundary between the normal state and the necessarily imperfect hypothetical lesion signal distribution. The choice of this distribution may thus introduce precisely the kind of bias the automated method is brought in to counter. One solution to this problem is therefore to be agnostic of the specific properties of the lesion signal, and identify everything that is anomalous in relation to the normal reference (Prastawa et al., 2004; Shen et al., 2010). Thus the pathology of a region is determined by its

remoteness from the normal case rather than its proximity to any hypothetical pathological case.

Fourth, whether or not the signal at any given locus is interpreted as normal or damaged often depends on the signal in its immediate anatomical vicinity.

Fifth, the optimal properties of an image on which to perform lesion segmentation are opposite to those of an image on which to perform lesion registration: this is so because in the former normal tissue contrast interferes with the lesion contrast one needs to distinguish normal from damaged brain, whereas in the latter lesion contrast interferes with the normal tissue contrast one needs to determine the anatomical labels of the lesion.

One solution, discussed in §1.3.3.2.2, is to use an anomaly score which indexes the anomaly of an unknown test datum either in relation to a reference set of data already known to be normal, or the rest of the data (Anbeek et al., 2005). Relating this to lesion segmentation, a test image is divided up into voxels with each one compared in turn to the homologous *k* instances within the reference set that resembles it most closely: its *k* nearest neighbours (Cover and Hart, 1967). Where the datum is a single value, a scalar, this is simply a matter of finding the *k* points that are closest to it on a linear scale. Where the datum has *n* variables describing it, a vector, this is some distance measure in *n* dimensional space, most simply the Euclidean, of the datum to its *k* nearest neighbours (*kNN*). Rather than discard the data within the clique of the *k* nearest neighbours one can take the mean of the distances of the test datum to each of the *k* nearest neighbours, a measure known as gamma (Harmeling et al., 2006).

The attraction of gamma is that it is indifferent to the shape and number of modes of the reference population distribution, and it is determined only by the reference set, requiring no prior knowledge of the properties of anomalous

data. The problem with gamma, however, is that it is sensitive to variations in the density of the reference population, that is, the degree of similarity between normal points within the reference set. This is illustrated in figure 1.5a, where a heterogeneous synthetic two-dimensional dataset is labelled by the gamma score of each point: the diameter of each point is drawn in proportion to the score. It is easy to see that members of the smaller, denser cluster will generally have a smaller value of gamma than members of the larger, sparser cluster even though each belongs equally strongly to its respective cluster. Thus, if anomaly were to be determined by a fixed threshold of gamma, either too many members of the sparse cluster will be labelled as anomalous or too few of the denser one. It is thus easy to see that the score is confounded by local differences in density.

So for gamma to be useful we need to find a way of correcting for such differences. As originally proposed in the context of computer network intrusion detection (Rieck and Laskov, 2007), an elegant way of doing this is to correct it by the mean of the distances between the members of the "clique" of k nearest neighbours, effectively a measure of local density. Thus, where the k nearest neighbours are sparsely distributed, gamma will have to be larger to be judged to be anomalous, and vice versa. The same dataset relabelled by this corrected score, zeta (ζ), is shown in figure 1.5b. We can now see that differences in local density no longer affect the score to the same degree. In short, the zeta score embodies the simple intuition that you are an outsider if those closest to you are closer to each other than they are to you.

In the present chapter the zeta anomaly score is applied to the task of lesion segmentation, calculating the zeta score independently for each voxel within a test image that has been brought into registration with a large set of images known to be free of lesions. The test datum is thus the signal at each voxel, and the reference data from which we choose the *k* nearest neighbours is the signal in each of the voxels at the same anatomical location across the reference set. Rather than using just the signal at a given voxel, a scalar value, the signal of the 26 immediately adjacent voxels (a 3 x 3 x 3 cube centred on each voxel) are extracted, resulting in a vector of 27 values for each voxel. This will provide information about local variations in the signal at any given point, a “searchlight” approach analogously used in multivariate functional imaging (Kriegeskorte et al., 2006). The result is a map of the brain where the anomaly of each voxel is labelled by its zeta score; appropriately thresholded this yields a binary lesion map of the brain.

2.2 Methods

All image processing and calculations were performed on a dual processor Intel Xeon 5600 2.4 GHz processor, with 24GB RAM and 120GB solid state drive (SSD). The raw dicom images were reconstructed into a 3 dimensional volume with xmedcon (Nolf et al., 2003), with all subsequent calculations performed in the MATLAB environment (The MathsWorks Inc.).

2.2.1 Imaging

A total of 208 brain scans were collected from the National Hospital for Neurology and Neurosurgery. All images were performed on a GE Genesis Signa 1.5 Tesla (1.5T) MRI scanner in a single session for each patient. The scans collected were the axially acquired T2-diffusion weighted echoplanar images (b0 and b1000 sequences) sampled at 1mm x 1mm x 6.5mm (TR=10000ms, TE=104.9ms, Columns=256, Rows=256, Slice thickness=5mm).

2.2.1.1 Focally lesioned brains

38 out of the total 208 brain scans contained focal lesions. These brain scans were unselected except for the presence of a visible vascular lesion on diffusion weighted imaging (DWI). The minimum lesion volume was 216mm^3. The mean age was 62.3 years (standard deviation (SD) = 17.8 years), and equal sex ratio. The mean time period between symptom onset and scan was 4.1 days (SD = 3.8) with a range of 1 - 17 days. This is slightly broader than the conventional notion of "acute". Consequently for some patients, the lesion will therefore have been visible on the b0 image used in the normalisation step, where the abnormal signal may have degraded the quality of registration to some degree as previously discussed. To the extent to which the quality of

normalisation impacts on the performance of lesion segmentation, better result might be achievable with a set of scans acquired only in the hyper-acute stage.

2.2.1.2 Non-lesioned reference brains

170 out of the 208 brain scans did not possess a focal brain lesion. A reference set of 95 non-lesioned brain scans with characteristics otherwise similar to the focally lesioned set was collected. The mean age was 67 years (SD = 15), and a sex ratio of 1.8 males to female.

2.2.1.3 Non-lesioned recipient brains

A separate set of 75 control subjects were matched to the lesioned brains by being drawn from a similar clinical population: patients attending a transient ischaemic attack (TIA) clinic who had not been found to have any acute lesions or any chronic lesions large enough to be visible on DWI. This was to facilitate the creation of a chimeric set. The mean age was 65 years (SD = 16), and a sex ratio of 1.3 males to female.

2.2.2 Image preprocessing

For each patient the corresponding b0 and b1000 sequences were reconstructed into a 3 dimensional volume using xmedcon (Nolf et al. 2003). The b1000 volume was rigidly co-registered to the b0 volume using the co-registration function in SPM5 (http://www.fil.ion.ucl.ac.uk/spm/) and its default parameter settings (the specific parameter settings are listed in appendix B).

2.2.2.1 Normalisation

The computation of zeta and any other voxel-wise statistic that compares a test image against a reference set of images implies knowledge of the

spatial correspondence between the two. In other words, the computation must be preceded by an inter-subject co-registration step in which spatially homologous voxels in the test and reference brain images are brought into alignment: a process conventionally referred to as normalisation (Ashburner and Friston, 1997, 1999; Brett et al., 2001; Friston et al., 1995). As discussed in §1.4.1.3 and §3.1, normalisation is commonly done by finding a set of linear and non-linear transformations of the target image such that differences in the signal at each voxel between the target and reference images are minimized at the cost of some plausible level of distortion. In the presence of a lesion, normalisation can therefore be difficult since the signal within the lesioned region will inevitably tend to interfere with the signal matching process (Shen et al., 2007).

One can minimize this effect by obtaining more than one imaging sequence and performing the normalisation on the sequence that is least sensitive to the lesion, as is performed here. Alternatively or additionally, one can mask out the lesioned region (Brett et al., 2001) or fill it in with signal from the homologous region in the contralateral hemisphere (Nachev et al., 2008), but this of course implies knowing the spatial parameters of the lesion. Another option is to combine the normalization and tissue segmentation (white and grey matter, and cerebrospinal fluid) into a single generative model, explicitly modelling the lesioned area as falling within a third, abnormal class. This procedure may be used iteratively, with the lesioned class derived from one run serving as a prior for the next (Seghier et al., 2008). Since the departure of the signal at each voxel from the expected value for grey and white matter respectively determines the extent to which it is likely to disturb the matching process, this approach would seem excellent at minimizing the impact of the lesion. As a means of lesion segmentation, however, it can only be expected to be satisfactory in the special case where the optimal normalization modality and the optimal lesion identification modality are the same. Clearly this is neither

inevitable nor desirable since the more the lesion disturbs the tissue contrast the worse the normalisation, and the greater the tissue contrast the more likely it is to interfere with lesion segmentation as already argued.

The approach here is therefore to derive the normalisation parameters by applying the unified normalisation and segmentation procedure implemented in Statistical Parametric Mapping 5 (http://www.fil.ion.ucl.ac.uk/spm/) to the b0 image, which has low lesion contrast and good grey/white matter contrast, and then to apply the resultant normalisation parameters to the corresponding b1000 image, thereby allowing much better normalisation than could be achieved from the b1000 image alone. Note that the success of the normalisation procedure is not dependent on the b0 image's being free of any lesion contrast because SPM5s routine will automatically eliminate most such voxels during the segmentation component as already explained (Seghier et al., 2008): it is simply desirable that this is kept to a minimum. We have previously shown that, in keeping with their T2-weighting, clinical b0 images of this resolution can be satisfactorily normalised (Nachev et al., 2008). This step is only practicable if the b0 and b1000 images are already in registration. Since each pair is taken in a single run, the images are bound to be close.

The b0 volume was spatially normalised into MNI space (ICBM space – European brains – template) using the combined segmentation-normalisation method in SPM5 (Crinion et al., 2007), with the derived normalisation parameters applied to the co-registered b1000 sequence. Again the default parameter settings for the segmentation-normalisation method provided by SPM were used, except for the interpolation setting in the writing option being set to a 7^{th} degree spline. All further operations were performed on the normalised images, which were resliced to 2 x 2 x 2 mm voxels.

2.2.2.2 Manual segmentation

In order to evaluate our segmentation algorithm we require a “ground truth” or a standard that defines what is truly lesioned. Since even if we could have a histological reference the post-mortem specimen would be anatomically too distorted to be usable in mapping, the only standard here can be manual segmentation by eye: this is the approach every other major study in the field has taken (Andersen et al., 2010; Bhanu Prakash et al., 2008; Brett et al., 2001; Gupta et al., 2008; Hevia-Montiel et al., 2007). Each lesion was therefore segmented by hand with the aid of MIPAV’s gradient, magnitude and direction live-wire edge detection tool (Barrett and Mortensen, 1997; Chodorowski et al., 2005; Falcao et al., 1998). The manual segmentation was performed on normalised versions of each image. An overlap of the images is shown in figure 2.1.

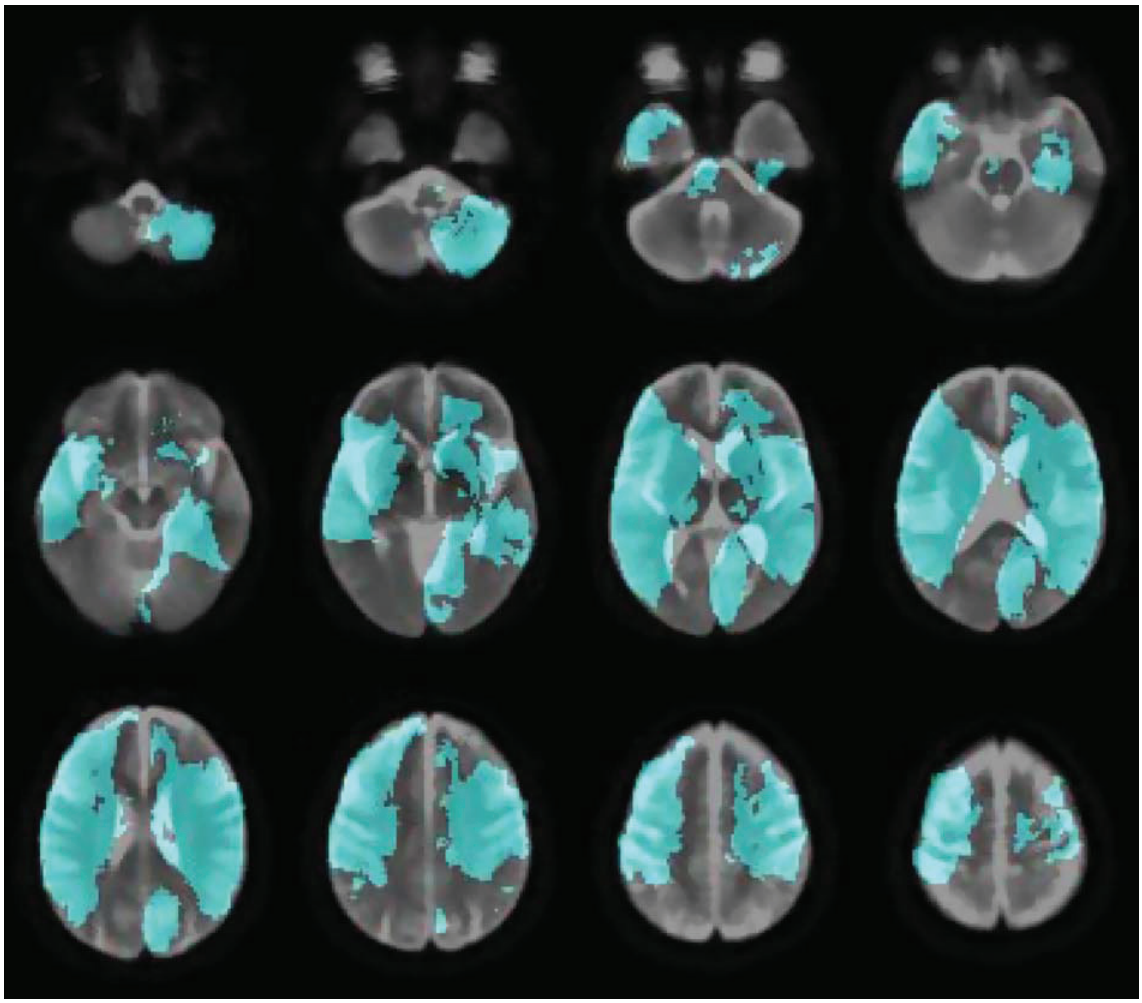


Figure 2.1 - Lesion coverage map.

The overlay (cyan) shows the coverage of the 38 manually segmented lesions. The underlay is the mean image created from the reference b0 dataset (n=95).

2.2.2.3 Chimeric image creation

The set of chimeric images (artificially lesioned brain images) was generated by replacing in each non-lesioned image and sequence (b0 and b1000) the signal intensities in the corresponding voxels within each of the lesions, resulting in 38 x 75 = 2850 chimeric images.

The point of this manoeuvre was to evaluate the algorithm's ability to distinguish lesioned from non-lesioned signal *at the same location.* Unless one is fortunate enough to have a pre- and post-lesion scan for several patients, which we are not, the only way of performing such an analysis is by using "chimeric" images of this kind (Nachev et al., 2008). If we had no grounds to suspect that lesion discriminability varied with location an analysis of this kind would have been unnecessary. However, it is clear that this cannot be assumed to be so, especially with DWI where some locations (e.g., frontal and temporal poles) are much more prone to artifactually abnormal signal than others.

For the chimeric images to be realistic the non-lesioned tissue in each donor/recipient pair must be broadly similar. Since, unlike CT scans, MRI images do not have a standardized intensity scale (Bergeest and Jager, 2008; Nyul and Udupa, 1999), each transplantation step must be preceded by equating the global intensity of the donor and recipient images. This was done by calculating the mean whole brain signal, excluding voxels falling within the lesion, of the lesioned, donor image and the mean whole brain signal of the recipient image, and multiplying the signal value at each voxel within the latter image by the ratio of the two means (Brett et al., 2001). This procedure ensured that the transplanted signal was as close as possible to what abnormal signal might have looked like had it been present in the recipient; without it, unrealistic global variations in the signal would have clouded our assessment of the algorithm's performance. We confirmed that this procedure did not artificially introduce an artefactual difference between the lesion signal

and the signal in the immediate vicinity of the lesion by comparing the ratios between intralesion (5 voxels inside the lesion boundary) and perilesion (5 voxels outside the lesion boundary) signal for all original lesion images and their corresponding sets of chimerics (paired t test p value < 0.01 for each set). Figure 2.2 is a flow diagram illustrating the process of creating these images.

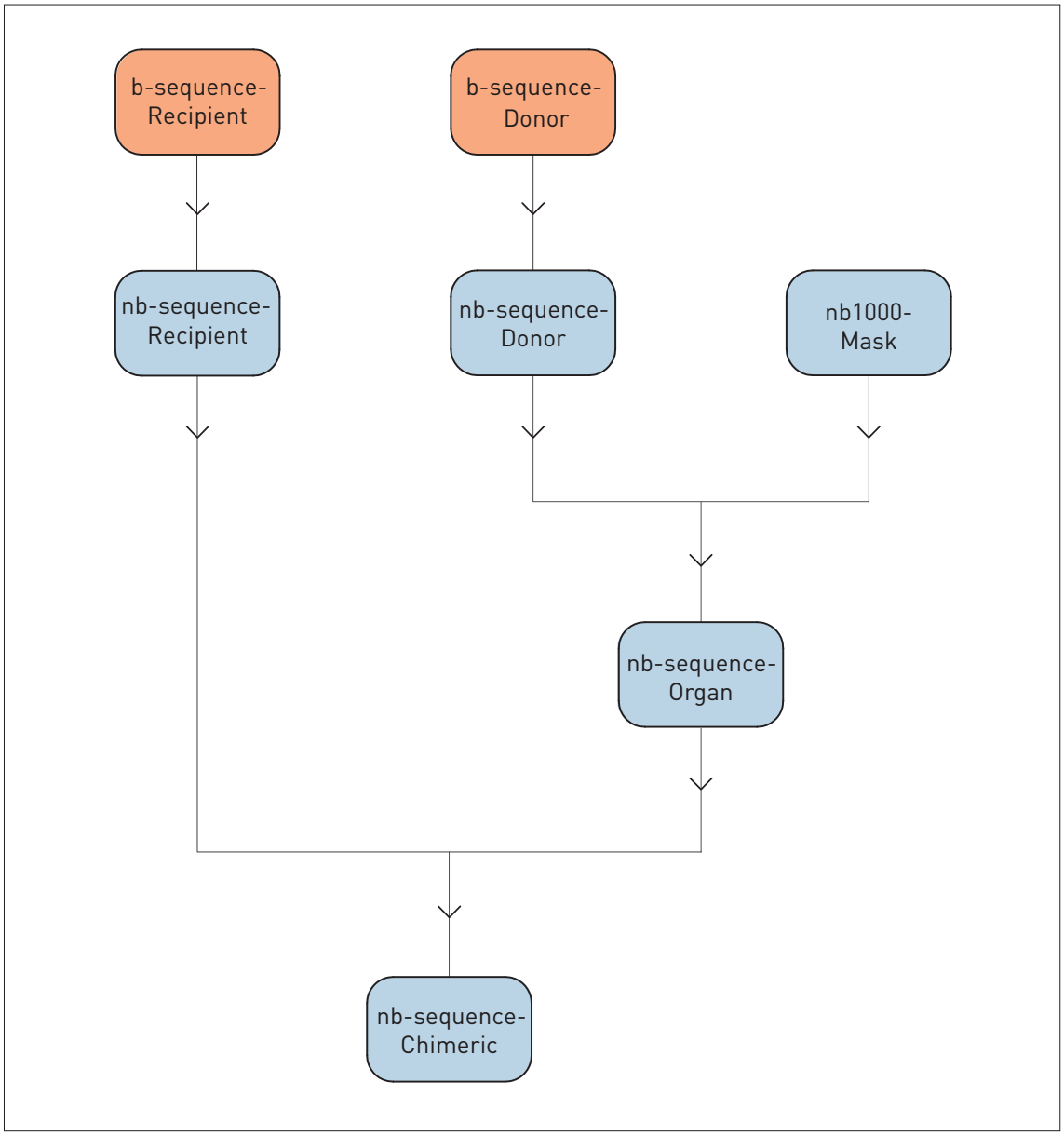


Figure 2.2 - Chimeric image creation flow diagram.

Both the recipient and donor images (b0 and b1000 sequences) were spatially normalised into standard MNI space using the unified segmentation-normalisation routine in SPM5 to create nb-sequence-Recipient and nb-sequence-Donor. The default settings for the unified segmentation-normalisation routine were used (appendix B), except for the interpolation being set to 7 in the normalise estimate and write parameters. The lesion within the donor b1000 sequence was then manually segmented. Whilst in MNI space, the binary mask was applied to the normalised b sequences (b0 and b1000) donor brain (nb sequence-Donor) to extract the lesioned voxels (nb-sequence-Organ), which was subsequently transplanted into the corresponding normalised b sequence recipient brain (nb-sequence-Recipient) to create a chimeric brain (nb-sequence-Chimeric).

Abbreviation	Description
T1-Recipient	Unlesioned T1 image in native T1 space
nT1-Recipient	Unlesioned T1 image in MNI space
b1000-Donor	Lesioned b1000 image in native b1000 space
nb1000-Donor	Lesioned b1000 image in MNI space
nb1000-Mask	Binary lesion mask in MNI space
nb1000-Organ	Extracted lesioned voxels in MNI space
nT1nb1000-Chimeric	Chimeric image in MNI space

Figure 2.3 - Table of abbreviations for figure 2.2.

The various images are named with the volume space first and their role second. All images are diffusion weighted scans.

2.2.3 Zeta (ζ) anomaly score

We have seen that the gamma (γ) score of a given point, *x*, is the average distance to its *k* nearest neighbours (nn)

$$\gamma_\kappa(x) = \frac{1}{k}\sum_{i=1}^{k} d(x, nn_i(x))$$

The zeta score is the difference between gamma, an index of the anomaly of the point in relation to its neighbours, and the average inner-clique distance of its neighbours, an index of the density of the neighbourhood clique (Rieck and Laskov, 2006).

$$\zeta_\kappa(x) = \frac{1}{k}\sum_{i=1}^{k} d\big(x, nn_i(x)\big) - \frac{1}{k(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{k} d(nn_i(x), nn_j(x))$$

To use the zeta score in the imaging domain, we need to make two key decisions. First is the value of *k*, the size of the neighbourhood clique, and the only adjustable parameter of the zeta score. The optimal value of *k* is determined by the smallest scale of any heterogeneity in the reference population, effectively the size of the smallest distinct cluster of values differing from the main population yet still within the normal range. Since the clustering of data has no established measure no rigid guide to the selection of *k* can be given. Here we arbitrarily set the value of *k* at 15% of the size of the number of cases within the reference set (15 images in our case).

The second decision to take is how best to parameterise the signal at each voxel: since we do not need to use a scalar, we can have a vector of values; we can incorporate as few or as many parameters per voxel as is computationally practicable. In this case we add information about the signal of adjacent voxels, thereby incorporating local pattern anatomical information

in the parameterisation of the signal at each voxel. The distance measure we compute at each voxel here is therefore the Euclidean distance in 27 dimensions, the signal at the index voxel and at the 26 voxels (a 3 x 3 x 3 cube) immediately adjacent to it, between each voxel in the test image and the homologous voxels and their adjacents in a set of reference, normal images co-registered with the index image so that anatomically homologous regions are brought into alignment. The result is a voxel-wise map of zeta values from which a binary image (lesioned vs non-lesioned) can be created by choosing an appropriate threshold zeta value. Thus though we use only one sequence to derive zeta, the b1000 image, we parameterise each voxel as the multivariate pattern of a 27 voxel cluster.

2.2.4 Zeta map thresholding

A zeta indexed map is a real numbered and dimensionless. It can be used directly to probe lesion-function relationships, however in order to assess its performance against the gold standard of manual segmentation, which necessarily generates a binary mask, an appropriate threshold must be found. For the algorithm to be unsupervised the threshold needs to be set automatically.

First, a whole brain map of the reference set data only was created. This diversity map would describe the variance in the normal signal across different parts of the brain. This was done by iteratively calculating the zeta value of each voxel within each brain image within the reference set only, taking the remainder of the set as the reference, in essence treating each member of the set as a test volume with the remainder being the reference. This produced a set of 95 whole brain zeta maps, one map per image in the reference set, giving us a set of empirical distributions of zeta values at each voxel within the reference set. These distributions were readily parameterised as generalized

extreme value distributions, with parameters location (μ), scale (σ), and shape (ξ) fitted using maximum likelihood estimation.

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\left(-\frac{1}{\xi}\right)-1}$$

This diversity map allowed us to index the degree to which the normal data at each anatomical location would exhibit the characteristics of an outlier, thereby allowing better discrimination of true outliers. We thus used the diversity map to correct each test zeta image by subtracting it from the median of the diversity map, calculated independently for each voxel location.

Second, this corrected zeta map was thresholded at 0, for any negative value would imply the signal at a voxel being less anomalous than that of any within the normal reference set.

Third, this coarsely binarized map was then submitted to a noise removal step where clusters with fewer than 5 voxels and/or fewer than 2 voxels with a zeta score of 70% of the maximum zeta value across the brain were removed.

Fourth, the surviving clusters, taken one by one, were then subjected to a further thresholding procedure, where starting at 0 and incrementally ascending the function, the zeta value was identified at which the slope of the relation between zeta and the volume of the cluster identified as anomalous at that threshold became greater than -10 voxels/zeta (for clusters smaller than 15000 voxels) or -100 voxels/zeta (for clusters larger than 15000 voxels). This cluster level zeta threshold was derived and applied independently for each cluster.

Fifth, the denoising step above was repeated.

Finally, a “hole filling” morphological operation was applied (since a completely disconnected island of normal tissue cannot possibly be functional) and

linear slice artefacts were automatically removed (based on their wholly non-anatomical spatial and signal characteristics), resulting in a final binary lesion image. The procedure is illustrated in figure 2.4.

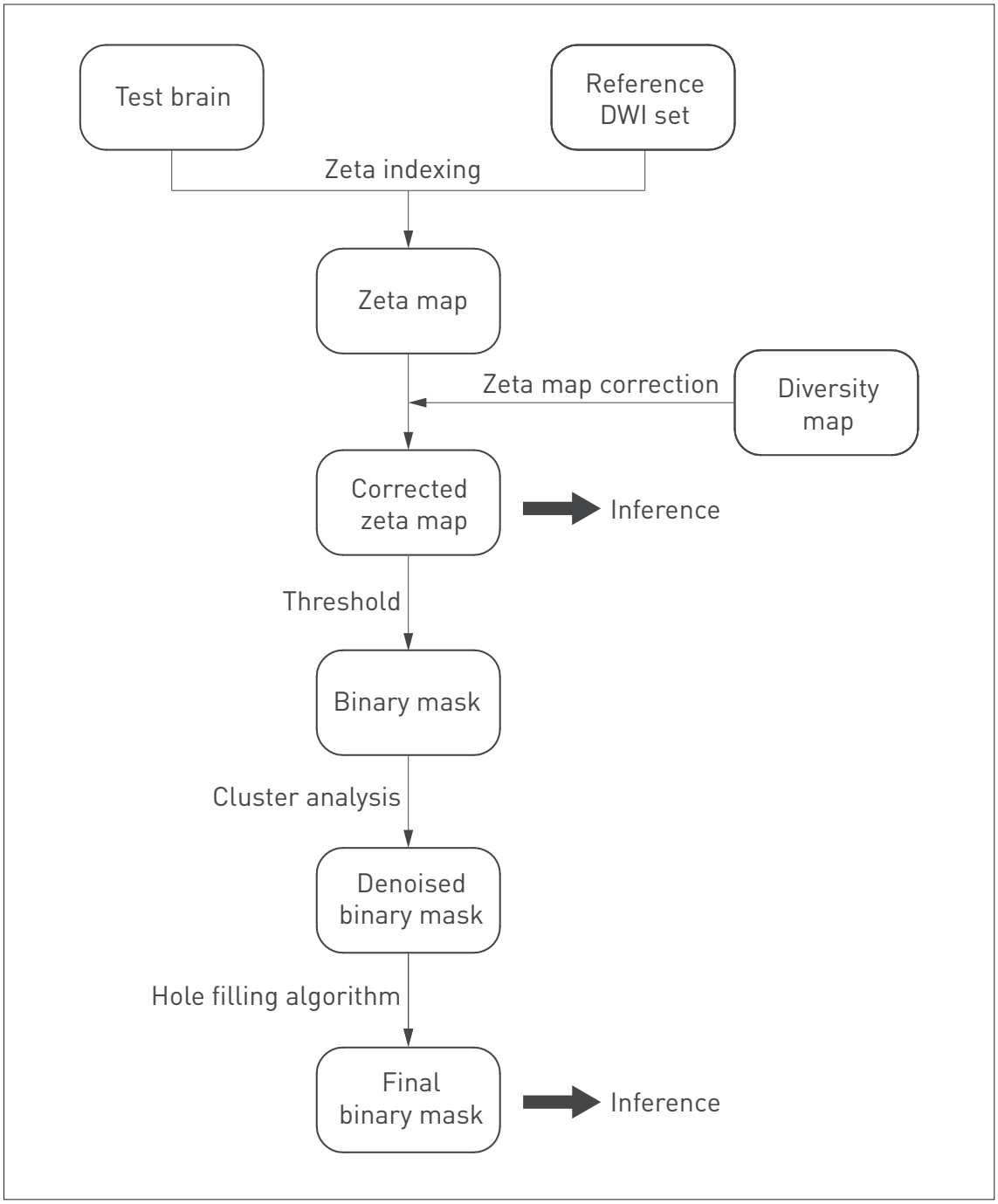


Figure 2.4 - Outline of zeta segmentation method.

A test brain is compared with a set of reference images in a voxel-wise manner to create a zeta map. This is then corrected using a diversity map that consists of the zeta maps for each normal brain in the reference set. The corrected zeta map may be used directly for inference, however in out case it is first thresholded to 0 to create a binary mask. The resulting clusters are then assessed in turn and are either removed or modified using a heuristic. Finally a hole-filling step is applied, to create a binary mask.

2.2.5 Evaluation

The evaluation produces a set of measures for each lesion natively, and each lesion-subject combination chimerically. In the native case, the variability of the measures can be estimated by lesion only. In the chimeric case, the estimates of the variability of each measure can be partitioned either by lesion or by subject. In the lesion-wise analysis, there are 38 groups, corresponding to each lesion, with replications within each group corresponding to each subject. In the subject-wise analysis there are 75 groups, corresponding to each subject, with replications within each group corresponding to each lesion.

The standard against which we compare the unsupervised segmentation is the manual tracing of each lesion by a trained operator. This is considered the "gold standard" in the field. The method of comparison follows established practice, (Anbeek et al., 2005; Dice, 1945; Shen et al., 2008; Zijdenbos et al., 1994) and use the following summary measures:

$$sensitivity = \frac{TP}{TP + FN} = \frac{voxels_{total} \cap voxels_{\zeta}}{voxels_{manual}}$$

$$specificity = \frac{TN}{TN + FP} = \frac{(voxels_{total} - voxels_{manual}) \cap (voxels_{total} - voxels_{\zeta})}{voxels_{total} - voxels_{manual}}$$

$$similarity\ index\ (SI) = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2(voxels_{\zeta} \cap voxels_{manual})}{voxel_{\zeta} + voxel_{manual}}$$

TP = true positive; *FP* = false positive; *TN* = true negative; *FN* = false negative

True positives are voxels correctly identified as lesioned, true negatives are voxels correctly identified as healthy, false positives are voxels incorrectly identified as lesioned, and false negatives are voxels incorrectly identified as healthy.

2.3 Results

The total computation time required to process a test brain image (spatially normalise the brain volume, create a zeta indexed map and segment into a binary volume), was in the order of 8 min.

2.3.1 Evaluation of the 38 native lesions

In the native case, the variability of the measures can be estimated by lesion only. This analysis shows a median sensitivity of 0.9602, specificity of 0.9979 and similarity index (SI) of 0.7342, with corresponding standard error of the median of 0.0074, 0.0007 and 0.0145 respectively (figure 2.5).

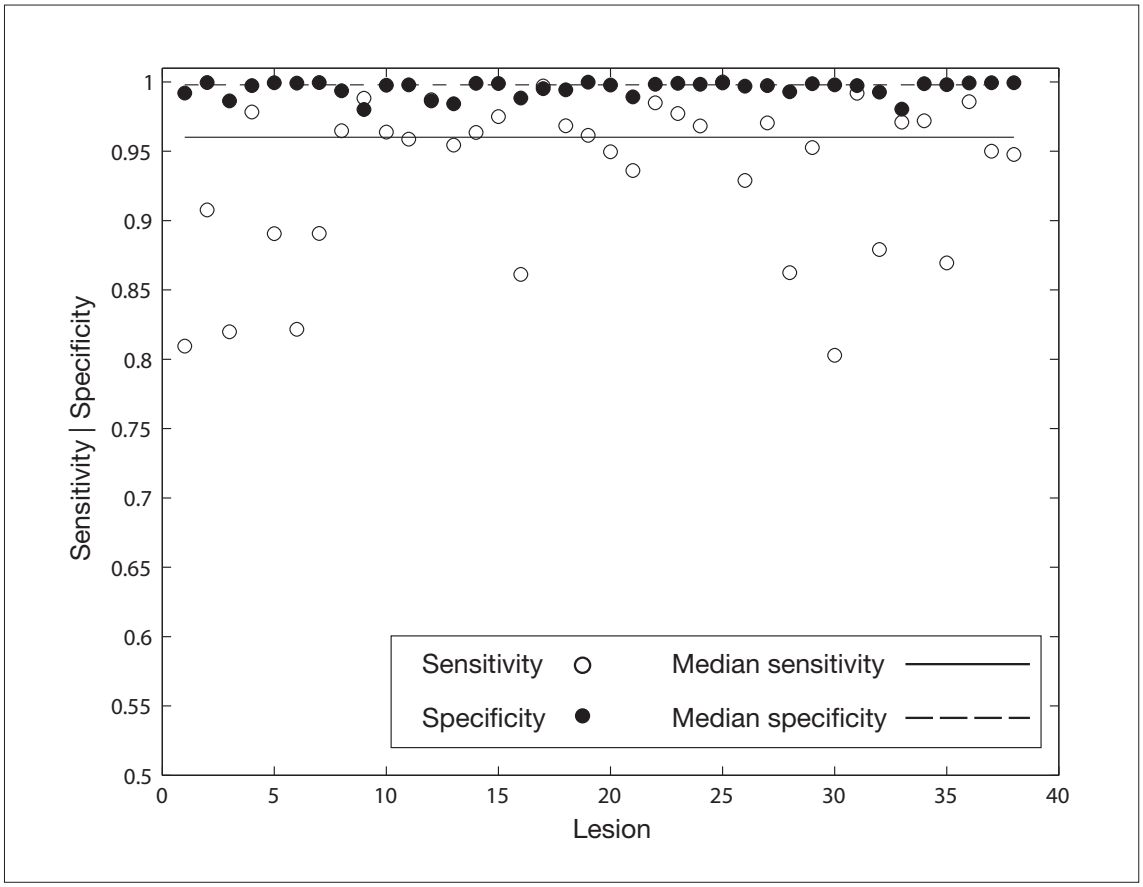


Figure 2.5 - Native brain performance.

Plots of sensitivity and specificity for each native lesion. A zeta map was created for each lesion image and a thresholded version was compared against the ground truth defined by manual segmentation. Note that the median sensitivity (dotted line) and specificity (solid line) for the 38 native lesions are high at 0.9602 and 0.9979 respectively.

2.3.2 Evaluation of the 2850 chimeric lesions

This analysis shows an overall median for sensitivity of 0.9362, specificity of 0.9986 and SI of 0.7840.

2.3.2.1 Chimeric images partitioned by lesion

In the lesion-wise analysis, there are 38 groups, corresponding to each lesion, with replications within each group corresponding to each subject. As illustrated in figure 2.7, the range of medians for the 38 groups in the lesion-wise analysis was 0.7578 – 1.000, 0.9807 – 0.9999 and 0.5724 – 0.9281 for sensitivity, specificity and SI respectively. The median sensitivity, specificity and SI for the above range of values were 0.9400 (SE = 0.0069), 0.9988 (SE = 0.0002) and 0.7829 (SE = 0.0118) respectively. This narrow variability suggests a robustness of the method to variations in lesion spatial and signal characteristics.

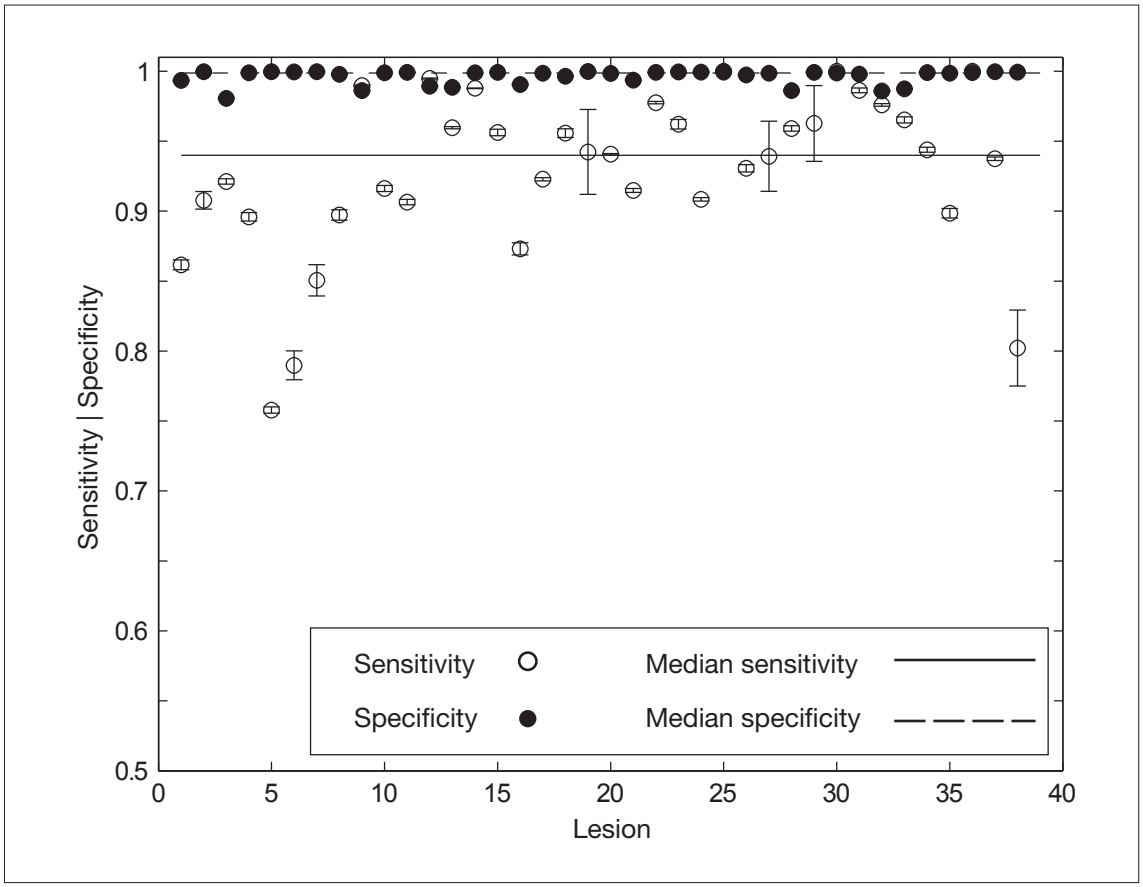


Figure 2.6 - Chimeric brain performance partitioned by lesion.

Plots of median sensitivity and specificity for each lesion calculated using all 2850 chimeric lesions, with each set containing 75 brain volumes. The error bars are standard errors of the median. Note that both sensitivity and specificity vary very little across the 38 lesions.

2.3.2.2 Chimeric images partitioned by subject

In the subject-wise analysis there are 75 groups, corresponding to each subject, with replications within each group corresponding to each lesion. The subject-wise assessment showed median values of 0.9399, 0.9988 and 0.7927, with standard errors of the median now 0.0049, 0.0001 and 0.0084 for sensitivity, specificity and SI respectively. Figure 2.7 shows the median values for the test parameters, across all lesions for each of the 75 recipient subjects. The minimal variability in the values, sensitivity 0.8296 - 0.9601, specificity 0.9899 - 0.9992, SI 0.5252 - 0.8102, suggests a resistance to noise and artefact.

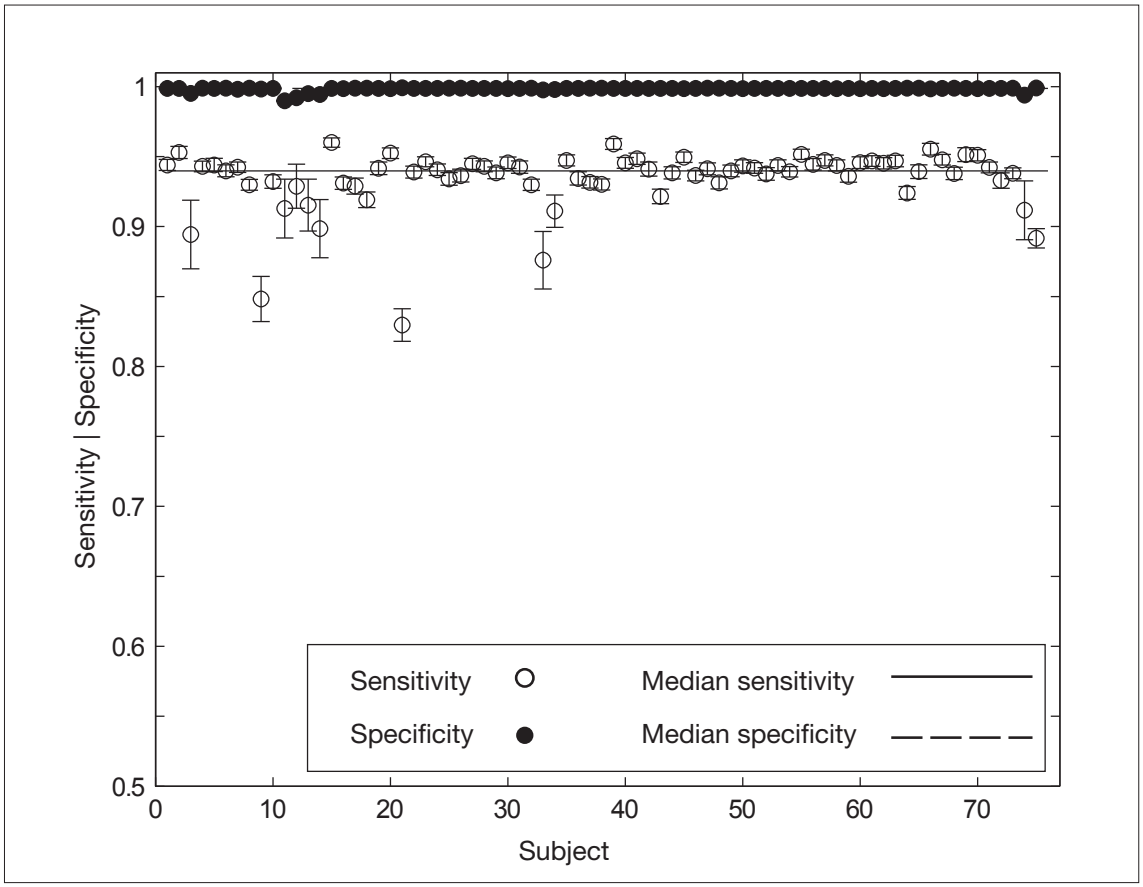


Figure 2.7 - Chimeric brain performance partitioned by subject.

Plots of median sensitivity and specificity for each lesion calculated using all 2850 chimeric lesions, with each set containing 38 brain volumes. The error bars are standard errors of the median. Note that both sensitivity and specificity vary very little across the 38 lesions illustrating the robustness of the method to variations in normal signal and artefact.

2.3.2.3 Visual comparison between the manual segmentation and unsupervised segmentation result for a cortical and sub-cortical example lesion

Figure 2.8 and 2.9 show a slice through an original b1000 native scan, with the associated zeta map, manual segmentation and segmentation result from the algorithm for a cortical and subcortical lesion respectively. On visual inspection, the similarity between the manual segmentation and the zeta segmentation result is very good. Importantly, the algorithm is able to appropriately ignore the high signal artefact at the temporal poles whilst still detecting the lesion (figure 2.9).

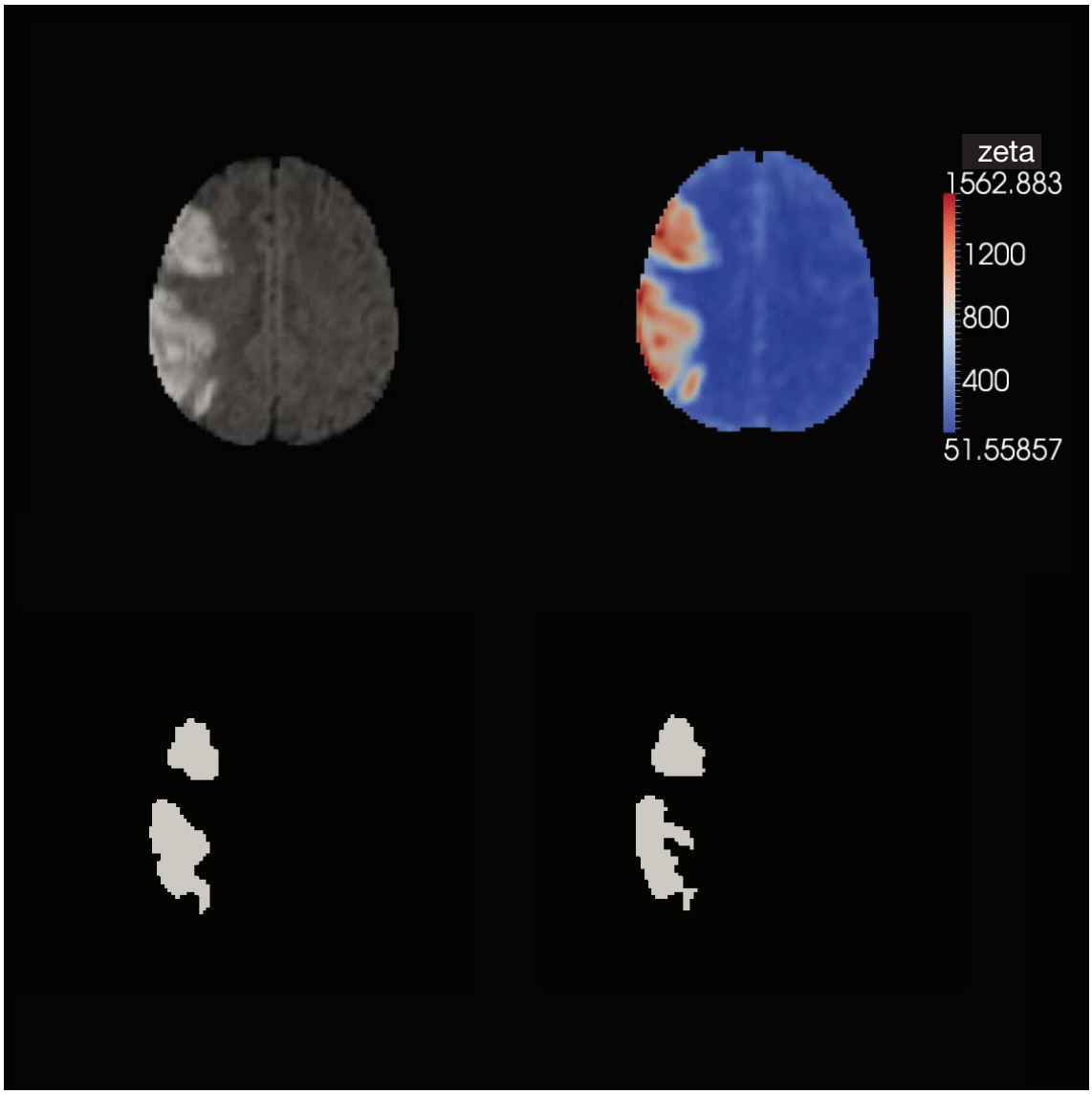


Figure 2.8 - Zeta segmentation (cortical).

All four images show the same slice from the normalised DWI brain volume of a patient with a large cortical stroke. Diffusion weighted image (top left). Zeta map (top right). Manual segmentation (bottom left). Thresholded zeta map (bottom right).

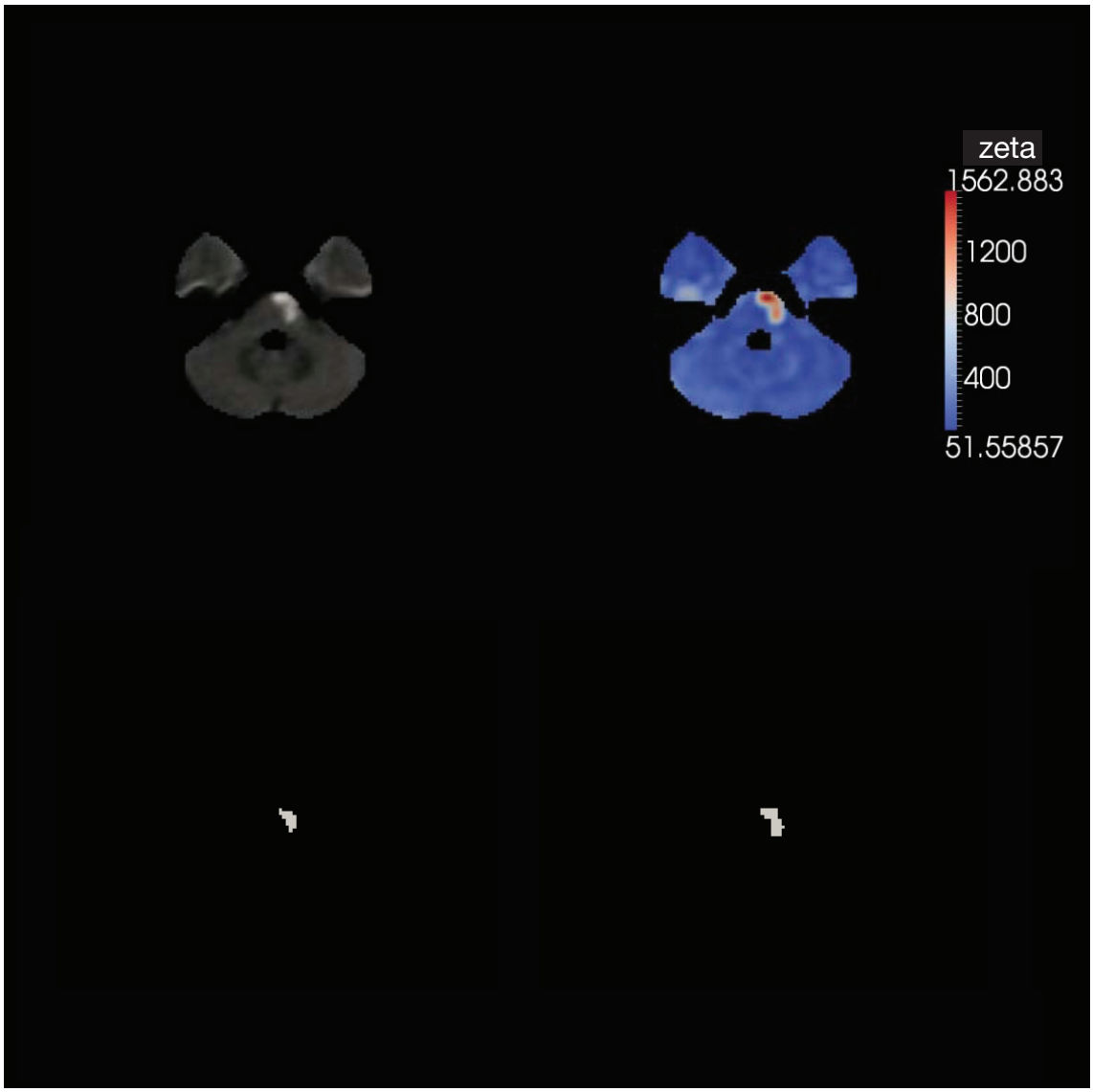


Figure 2.9 - Zeta segmentation (subcortical).

All four images show the same slice from the normalised DWI brain volume of a patient with a subcortical stroke. Diffusion weighted image (top left). Zeta map (top right). Manual segmentation (bottom left). Thresholded zeta map (bottom right). Note how the method appropriately ignores the regions of high signal artefact at the temporal poles.

2.4 Discussion

We have proposed a simple, general, unsupervised method for segmenting brain lesions derived from a recently described measure of anomaly and theoretically applicable where a reference set of normal data is available. Evaluation of its fidelity against manual segmentation in the context of diffusion weighted imaging of ischaemic lesions has shown good performance across a range of subjects and lesion parameters. The method is theoretically easy to adapt to other lesion types and imaging modalities, either alone, or in multispectral combination, and creates the possibility of high-throughput, fully-automated image processing pipelines for conducting large scale lesion-function studies. The following section discusses the advantages, and disadvantages of this new method.

2.4.1 Advantages

First, the performance metrics of zeta segmentation, at the time of assessment, exceed those of other published algorithms for diffusion weighted imaging with very close approximation to the current gold standard of manual segmentation (appendix A). There is little variability across subjects and lesions within which a bias could conceivably emerge, indeed less variability than one typically finds between observers (Fiez et al., 2000). Clearly, since manual segmentation is the gold standard here, it is not possible to conclude that zeta is better than manual segmentation though it may be more consistent.

Second, the naturally adaptive nature of the core anomaly metric theoretically allows zeta segmentation to handle highly heterogeneous data, adapting to any idiosyncrasy of a particular sequence, in our case with the prominent regional artefact seen on diffusion-weighted imaging. Zeta is able to distinguish artifactually high signal in such areas from true lesion signal elsewhere.

Third, zeta segmentation can be applied to any dataset without prior knowledge of the spatial or signal features of the lesions. It requires no priors for the number of lesions, their pattern of distribution, their signal intensity, or other features automated segmentation algorithms frequently demand from the user. Since lesions are both relatively rare, and so heterogeneous that their characteristics cannot easily be parameterised, it is advantageous that the method does not require information that is difficult or impossible to acquire.

Fourth, the high-dimensional nature of zeta makes it relatively easy to incorporate local anatomical information in the decision whether or not to label a voxel as abnormal by deriving the anomaly value not only from the signal at each voxel but also from voxels that are anatomically adjacent. The method thus becomes sensitive to variations in the local pattern of the signal, not just its point value, just as a human operator would naturally perform the task.

Fifth, zeta segmentation can easily be made multispectral: all that is required is for the different imaging modalities to be in spatial register. Since co-registering different modalities of a single subject is generally substantially easier than co-registering a single modality to a standard template this is simple to implement. For example, in the case of stroke it is possible to compensate for signal drop out in older lesions by adding another modality, such as a T2* sensitive sequence, in the distance matrix. The anomaly of each point will then be calculated on the basis of both sequences. Indeed, the only limits to the number of modalities one can add are computational, and these are likely to be modest.

Sixth, zeta segmentation naturally produces a real numbered image where the anomaly of each voxel is given a continuous index. Since there are rarely any physiological grounds for drawing a sharp threshold at which neural tissue ceases to be active this seems to us a more appropriate instrument than any

kind of binary measure. The zeta map can be directly used in making lesion-outcome inferences or relating patterns of damage to other factors of interest.

Seventh, zeta has only one parameter to adjust, the size of the clique, *k*. The lower the value of *k* the smaller the scale of signal inhomogeneity to which the algorithm is sensitive. We have not found a need to explore a wide range of *k*, but if this is necessary a univariate search would be quick and simple to implement. It is also theoretically possible to adjust *k* for each voxel independently in response to sequence-specific voxelwise variations in the pattern of inhomogeneity, perhaps because of distortion or artefact. If binary maps are required other parameters come into play depending on the chosen heuristic for thresholding; this is not necessary if the zeta maps are used directly, as we would suggest is the best approach. Finally, zeta is computationally economical, taking approximately 8 minutes per image running on commodity hardware.

2.4.2 Disadvantages

First, our method may be taken to assume that there is a monotonic relation between signal abnormality in a region and the probability or degree of dysfunction. Naturally, it may not be so, but until we have an independent means of establishing the link between physiology and the MRI signal in a given sequence, the signal intensity is all we have. Lesion segmentation is a segmentation of images, not of the brains they imperfectly reflect. This deficit is therefore common to all segmentation algorithms.

Second, the zeta anomaly score is a continuous variable with no *a priori* criterion on which one could discretize it. Where binary maps are required a heuristic method of selecting a threshold is necessary. However, the performance of the heuristic we have adopted here is close to that of a trained

operator, whose one criterion is by its very nature heuristic. Whether it would also do so for other modalities is a matter to be established empirically. Given the nature of the zeta score, it is to be expected that data with similar lesion contrast-to-noise ratios would perform similarly, but naturally this is not something that can be guaranteed. Although binarized data simplifies the business of inference, labelling every part of the brain as either completely lesioned or completely healthy, it involves assumptions that are neither justified empirically nor plausible *a priori*. While the CSF-filled centre of a chronic lesion clearly cannot have any function whatsoever, there is no reason to suppose that our arbitrary labelling of the margins of the lesion inevitably corresponds to a critical level of deterioration of physiological function, at least not for a great many lesions. Indeed, the habit of using discrete maps is arguably an artefact of the traditional way of segmenting lesions, by hand-drawn line. In any event it is a feature of the general approach to making inferences using lesions in the brain, not of this particular method.

Third, in common with any voxel-wise algorithm, zeta segmentation relies on the images already being spatially in register. Until recently, this was a major obstacle because without knowledge of the spatial characteristics of the lesion it is difficult to minimize its impact on the normalisation process. If the normalisation is imperfect, so necessarily will be any subsequent operation. However, SPM5's combined normalisation and segmentation routine has been shown successfully to overcome this problem (Andersen et al., 2010; Crinion et al., 2007). Critically, that study used images for normalisation with very strong lesion contrast, maximally testing the algorithm's capacity to deal with lesioned images, and so one would expect their findings to extend to other modalities. The improvement in performance is roughly four times that of the preceding gold standard (Brett et al., 2001), resulting in very small error values per voxel. One can further minimize the impact of this problem by deriving the normalisation parameters from a separate imaging sequence with low lesion

contrast, as we have done here, and as we suggest is the theoretically optimal approach.

Fourth, as with other voxel-wise methods, “islands” of normal tissue completely disconnected from the rest of the brain by damaged tissue will nonetheless be scored as normal. While in the special case of completely surrounded islands this is easily dealt with by the simple, hole-filling, morphological operation we employ here, it would be difficult to construct a method that deals robustly with intermediate cases. Once again, this is not something that any automated algorithm could easily solve without detailed knowledge of the connectivity of each brain area and the location of each connecting tract: information that we do not yet have to any degree of precision. Rather than making *ad hoc* decisions in each particular case, as is implicit in manual segmentation, it is perhaps best not to attempt an automated (or indeed manual) solution to this problem. Given that most lesions are more or less ellipsoidal, any distortion resulting from such effects is likely to be of relatively minor importance.

Fifth, zeta requires a set of normal images to use as a standard reference. While the relatively modest size employed here appears to be sufficient, this may not be so for other modalities and lesion types. Clearly, the performance of the algorithm will be dependent on how closely the standard set is matched to the test images: ideally, the only difference between them should be the presence or absence of lesions of the type being sought. These do not need to be of normal subjects but merely of patients without demonstrable acute ischaemic lesions. Indeed, it would be desirable to choose a reference cohort that is matched to the test population in all respects except the presence of acute ischaemic lesions, following the basic principles of “control” selection in experiments generally.

2.5 Conclusion

We have devised a simple unsupervised lesion segmentation algorithm based on zeta, a recently-described anomaly score. The algorithm places minimal demands on the experimenter and the data, and has wide potential applicability. Tested against a dataset of vascular lesions captured by DWI, it compares favourably against manual segmentation across a range of lesion sizes, locations and morphologies.

3 Optimal inter-subject registration of human magnetic resonance brain imaging in the presence of focal lesions

3.1 Introduction

To make spatial inferences across a group of brains, the images must first be brought into register with one another such that spatially homologous regions in the two scans are in alignment. This is usually achieved by using either a group mean average of the brains or a predefined standard template. This process of so called spatial normalisation is essential to ensure that only homologous brains regions are compared with each other, with the accuracy of the process influencing the level of noise carried over into the inferential stage. The state of the art methods used to perform this task are the Diffeomorphic Anatomical Registration Through Exponential Lie algebra (DARTEL) (Ashburner, 2007) and its predecessor, Unified Segmentation (Ashburner and Friston, 2000) implemented in the Statistical Parametric Mapping software (http://www.fil.ion.ucl.ac.uk/spm/).

The unified segmentation-normalisation routine was released in SPM5 and performs within the same iterative model, image segmentation, bias correction and spatial normalisation. It uses three template priors (grey matter, white matter and cerebrospinal fluid) to calculate the normalisation parameters. New segment is essentially the same as the unified segmentation-normalisation routine except that it utilises additional tissue priors (grey matter, white matter,

soft tissue, cerebrospinal fluid and air / background) with slightly different mixing proportions. It is not limited to the priors supplied with the software suite, but can accommodate as many extra templates as desired by the operator. DARTEL (Diffeomorphic Anatomical Registration Through Exponential Lie Algebra) is a framework that involves fitting a model to the data, where the model parameters encode the relative shapes of whatever the image represents in the form of flow fields. The function (derived model) that maps the transformations from native space into normalised space is invertible with both the forward and inverse function being smooth.

Both the Unified segmentation-normalisation routine and DARTEL are segmentation dependent. Unlike the unified segmentation-normalisation method, image segmentation is not part of the DARTEL routine. Consequently, it requires the segmented tissue classes (grey and white matter) to be provided. This can be achieved by using either the segmentation result from the unified segmentation normalisation routine or with New segment.

Although these methods have been shown to produce excellent results in normal healthy brains using T1 weighted magnetic resonance imaging (MRI) (Klein et al., 2009; Tahmasebi et al., 2009; Yassa and Stark, 2009), the task is complicated in the presence of a lesion. This is because these methods rely on minimizing the root mean squared error between the test and template images. Since there is no appropriate place in the template for injured tissue to reside – because by definition lesioned tissue is not normal – such volumes are essentially missing information.

Earlier work by Brett et al (Brett et al., 2001) proposed ignoring the lesion volume when calculating the normalisation parameters, in a process termed cost function masking (CFM). This manoeuvre prevents the aberrant lesion information from interfering with the normalisation process. Later Crinion et al (Crinion et al., 2007) suggested CFM was unnecessary when the

unified segmentation-normalisation routine was used with medium to high regularisation. This technique initially segments the source image (the image with the lesion) into different classes, with lesioned tissue placed into a separate class that is excluded from the model. In this way the technique implicitly performs cost function masking and avoids the need to manually specify the lesion boundary prior to normalisation. The problem that follows is lesioned tissue that happens to match the signal intensity of a neighbouring, but incorrect, tissue class will still be allowed to influence the result. In contrast to ignoring the lesioned volume, Nachev et al (2008) proposed replacing the injured voxels with the signal intensities of the corresponding healthy values from the contralesional hemisphere. By exploiting the significant mid-sagittal symmetry in the brain, substituting the lesion tissue for the healthy contralesional voxels would provide a more accurate representation of the brain prior to its injury as the intersubject differences are much larger than interhemispheric differences (Nachev et al., 2008). This still required the lesion border to be delineated before execution of the normalisation routine. Alternatively, an extension to the unified segmentation-normalisation routine was proposed by Seghier et al where a third and fourth additional tissue class is utilised for cerebrospinal fluid and voxels of “unknown” class that do not fulfil the grey and white matter criteria (Seghier et al., 2008), thus performing an automated cost function masking procedure.

Recently the debate over the benefit of CFM has been resurrected in a paper by Andersen et al suggesting that inclusion of the lesioned volume in the normalisation procedure results in compression of the lesion volume by the unified segmentation routine (Andersen et al., 2010). This volume reduction is measured as the ratio between the volumes derived from the method in question and the unified segmentation plus CFM routine. This phenomenon is not a uniform effect; rather larger lesions display a greater compression than

smaller ones, though this excessive warping could be reduced by CFM and the use of a coarse lesion mask.

With the return of the CFM debate, Ripollés et al (2012) have investigated the performance of large deformation routines like DARTEL on brains with vascular stroke lesions and brain resections. Interestingly the performance of the unified segmentation routine with CFM was comparable to DARTEL with CFM, while the latter still experienced a small effect of lesion volume compression. However when the authors examined the DARTEL routine in combination with their custom white matter mask (an additional tissue class prior template created by taking the mean of the white matter and CSF prior found in the SPM8 suite) applied in New segment there was a significant reduction in the root mean squared displacement and corresponding increase in the normalized cross-correlation score – a modified similarity index – compared with the unified segmentation routine. Unfortunately, this DARTEL combination produced a 17% reduction in lesion volume (Ripollés et al., 2012). In the end the authors conclude that depending on the metric used to assess the quality of the normalisation process, there is a trade-off between the optimal root mean squared displacement and lesion volume reduction.

Following Ripollés et al (2012) discussion, the optimal root mean squared displacement was from DARTEL in combination with CFM with a value in the order of 0.6mm. This was however with a relatively large reduction in lesion volume of 9.4%. The enantiomorphic normalisation process proposed by Nachev et al (2008) incorporates prior information about the brain by borrowing the signal intensities of the corresponding healthy voxels from the contralesional hemisphere. Their results revealed a root mean squared error in the order of 0.06mm. Though the method of artificial lesion creation was slightly different to the latter studies, the reduction in displacement is still comparatively large.

Any normalisation algorithm applied to lesioned brains is limited to the amount of information it has on what the brain looked like before the lesion. In the case of cost function masking, it attempts to limit the relevant information to areas that are not damaged. This is problematic for two reasons. First, it means that the quality of the registration will inevitably vary with the quantity of brain masked, in a way that can potentially cause a bias in any subsequent inferential analysis. Second, where CFM is implicit it will vary from lesion to lesion to the extent to which cost function masking happens to work. Neither of these scenarios is desirable.

Alternatively one can search for some surrogate of the signal intensities for the lesioned voxels. The brain exhibits significant symmetry along the mid-sagittal plane. Consequently signal intensities from the homologous region of undamaged contralesional hemisphere can provide an indication of the state of the brain prior to the injury. Crucially the inter-hemispheric differences within subjects can be expected to be lower than inter-subject differences (Thompson et al., 1996; Watkins et al., 2001). Moreover the error from the inter-hemispheric differences should asymptote with the that from the inter-subject differences as lesion size increases (Nachev et al., 2008).

Since the arrival of DARTEL and New segment, there have been a few studies reporting favourable spatial normalisation results in comparison with the unified segmentation-normalisation, cost function masking combination. However, no assessment of these new routines with enantiomorphic correction (Nachev et al., 2008) have been reported. In this chapter I shall address this question and examine the performance of these routines in the presence of a lesion with and without enantiomorphic correction.

3.2 Methods

In the case of a lesioned brain image, assessing the fidelity of inter-subject registration is not possible objectively as there is no obvious metric of how well it is accomplished. One cannot look at the error with the template, because an algorithm that has transformed an image into the template while preserving none of the features of the lesion would have a fidelity of 100%. Alternatively one can look at manually defined landmarks but such things will always be sparse, there are only so many you can place, and subjective in their placement on the source and evaluation of position on the template. The focus of this study is not how well the normalisation would be but how well the normalisation copes with the disruptive effect of a lesion. To achieve this we have to compare “the same brain”, with and without a lesion. Such datasets are very difficult to find. Therefore we have to create a synthetic dataset in which real lesions are introduced into real brains to create unreal “chimeras” as closely matching the before and after state as possible.

One further issue to raise is the complexity of the lesion space, i.e. high-dimensional correlations between voxels that are affected in a group of lesions. We show in §4.3.1 that this space is very complex, and so to assess the impact of lesions generally one needs much larger samples than are commonly used in such studies. This is why we take a fully automated approach to extraction of the anatomical labels of the lesion, creation of the chimeric brain image and subsequent assessment of the normalisation routine.

To assess the different spatial normalisation routines 2 sets of magnetic resonance imaging (MRI) brains scans are needed. The first set should contain a collection of images that do not possess any focal abnormality. Registration of these images, in their unadulterated state, to standard (MNI) space by a base method will then act as a “baseline” transformation result.

Next these images will then act as a recipient image, into which a lesion – organ – derived from the second set – donor brains – of scans that contain focal lesions which have been first registered independently into standard (MNI) space. The created chimeric images are then transformed back into the corresponding native recipient space, to ensure the same starting point is used for both images with and without a focal lesion (Brett et al., 2001; Nachev et al., 2008; Ripollés et al., 2012). Note close registration between the recipient and donor brain is not essential at this step as we wish to see how the spatial normalisation routines handles images with focal lesions. By comparing the normalisation result with and without a lesion we can then quantify the effect of perturbation caused by the lesion on each normalisation routine.

All simulations were performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) in the MATLAB environment (The MathsWorks Inc.) on a dual processor Intel Xeon 5600 2.4GHz processor, with 24GB RAM and a 120GB solid state drive (SSD).

3.2.1 Data

3.2.1.1 Recipient set

11 T1 magnetic resonance imaging (MRI) brain scans were acquired on a Siemens Avanto 1.5 Tesla MRI scanner with syngo MR B15 software. The scans collected were the T1-weighted spin echo images (TR=430ms, TE=11ms, Columns=512, Rows=512, Slice thickness=5mm). All images were unselected except for the absence of any focal abnormality and collected from the National Hospital for Neurology and Neurosurgery, London (NHNN).

3.2.1.2 Donor set

The b0 and b1000 sequences for 77 T2 diffusion weighted MRI brain scans were collected. They were acquired on a Siemens Avanto 1.5 Tesla MRI scanner with syngo MR B15 software (TR=3200ms, TE=81ms, Columns=128, Rows=128, Slice thickness=5mm). All images were unselected except for the presence of a focal abnormality and collected from NHNN.

3.2.2 Image preprocessing

3.2.2.1 Registration of focally lesioned images and binary mask creation

In order to create an artificially lesioned brain, the donor brain and its accompanying manually segmented binary mask must first be brought into register with its recipient brain. We can either transform the donor brain into the native space of the recipient brain, or, as we do here, transform the recipient brain and donor brain plus binary mask into standard MNI space. Importantly this decision will not affect the subsequent analysis as this is not where we evaluate any normalisation fidelity. The donor brain containing the signal intensities of lesion we wish to use is the diffusion weighted b1000 MRI scan which has been acquired in the same pass as the b0 sequence. The b1000 sequence is particularly useful for detecting vascular lesions, with very good contrast between lesion and normal tissue. Unfortunately, the grey-white matter contrast (tissue contrast) is as a consequence rather poor, rendering it less useful for spatial normalisation. The b0 sequence, however, has good tissue contrast which is sufficient for spatial normalisation (Nachev et al., 2008). Since the 2 sequences have been taken from the same patient and in the same pass, their registration will inevitably be very close. The b1000 can simply be rigid body aligned to the b0 sequence, and the spatial normalisation routine

applied to the b0 sequence with the derived parameters subsequently applied to the co-registered b1000 sequence and binary mask to bring both images into standard MNI space.

A focally lesioned b1000 diffusion weighted (DWI) MRI scan is first rigid body co-registered with its corresponding b0 weighted DWI scan using the co-registration routine and default settings in SPM8. Next a binary lesion mask is created from the b1000 sequence resulting in 3 images – b0, b1000 and binary mask – all in register. The b0 sequence is then spatially normalised into MNI space using the unified segmentation-normalisation routine in SPM8, with the derived parameters subsequently applied to the b1000 and binary lesion mask. Default settings were used (appendix B) except for the interpolation being set to 6 in the normalise estimate and write parameters.

3.2.2.2 Chimeric brain creation

To facilitate the evaluation of each normalisation routine, a series of chimeric images were created using a similar method to Nachev et al and Brett et al (Brett et al., 2001; Nachev et al., 2008), where the signal intensities of injured voxels were extracted from a lesioned brain and transplanted into their corresponding locations in a healthy recipient brain volume. Critical to this manoeuvre is maintaining the registration between the recipient healthy brain and the artificially lesioned chimeric brain, with the only difference between the two being the signal intensities of the voxels within the boundaries of the lesion mask. This is necessary to quantify the effect of the lesion on the normalisation procedure. Although the chimeric images are not truly physiological, it is our best alternative to collecting a cohort of pre and post lesioned brains. Figure 3.1 is a flow diagram illustrating the process of creating these two sets of images.

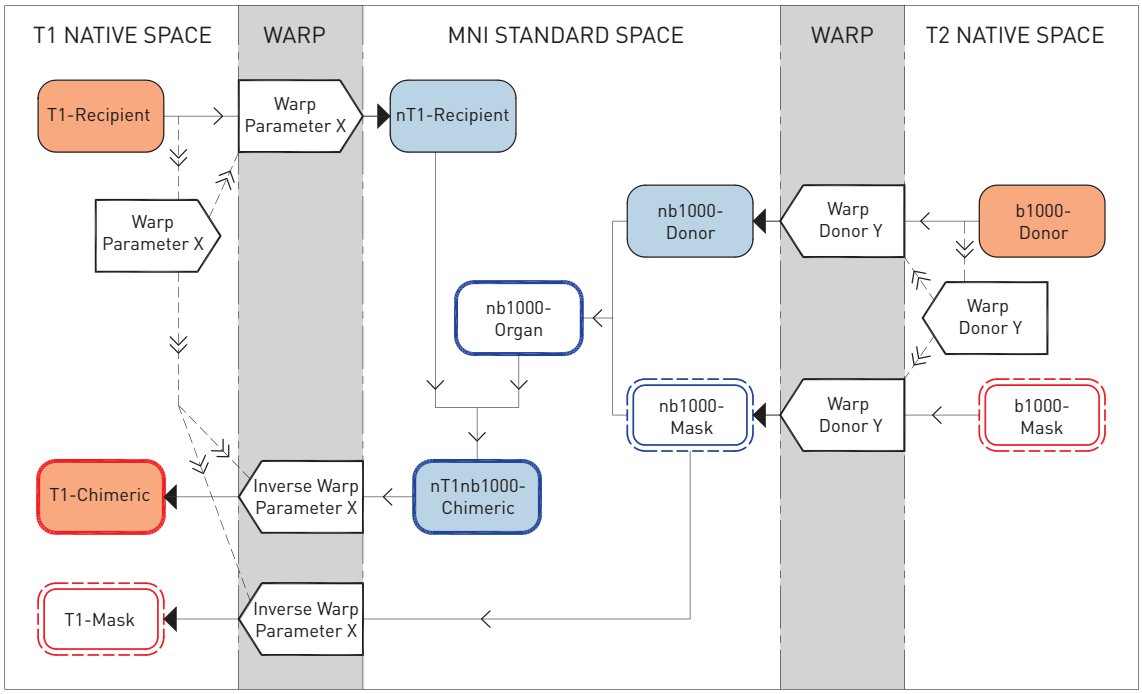


Figure 3.1 - Image pre-processing.

In order to transplant the lesion into the recipient brain it is convenient to move everything into a normalised space. It could be done directly (from one brain image to another, in each other's native spaces) but that is how we do it here. Crucially, a) this is not where we evaluate any normalisation fidelity and b) the quality of the normalisation is not so important because it only affects the similarity of the donor lesioned brain to the chimeric brain, which does not matter much as lesions vary from one brain to another a lot more than any noise this would introduce. The T1-Recipient brain volume is normalised into standard MNI space by using the unified segmentation-normalisation routine in SPM8. Similarly, the b1000 Donor volume and its complementary binary mask is first co-registered to its corresponding b0 volume. They are then transformed into MNI space using the normalisation parameters derived from the unified segmentation-normalisation routine performed on the b0 sequence. The normalised mask (nb1000-Mask) is used to extract the lesioned voxels from the normalised b1000-Donor volume, thus creating the "organ", which is subequently transplanted into the normalised T1 brain (nT1-Recipient). The resultant chimeric brain (nT1nb1000-Chimeric) is finally transformed back into native T1 space along with the nb1000-Mask, by applying the inverse warp parameters originally used to normalise the T1 brain into MNI space.

A binary mask of the lesion within the b1000 donor brain is created whilst in the native space. Next both the T1 recipient and b1000 donor scans are separately spatially normalised into MNI space using the unified segmentation-normalisation routine in SPM8 to create nT1-Recipient and nb1000-Donor.

The default settings for the unified segmentation-normalisation routine were used (appendix B), except for the interpolation being set to 6 in the normalise estimate and write parameters. The binary mask was then transformed into MNI space using the normalisation parameters derived from the b0 donor brain process.

Whilst in MNI space, the binary mask was applied to the normalised b1000 donor brain (nb1000-Donor) to extract the lesioned voxels (nb1000-Organ), which was subsequently transplanted into the normalised T1-recipient brain (nT1-Recipient) to create a chimeric brain (nT1nb1000-Chimeric). The chimeric brain along with the normalised binary mask (nb1000-Mask), were finally transformed back into the native space of the T1 brain by applying the inverse normalisation parameters derived from the T1-recipient brain process.

Abbreviation	Description
T1-Recipient	Unlesioned T1 image in native T1 space
nT1-Recipient	Unlesioned T1 image in MNI space
b1000-Donor	Lesioned b1000 image in native b1000 space
nb1000-Donor	Lesioned b1000 image in MNI space
b1000-Mask	Binary lesion mask in native b1000 space
nb1000-Mask	Binary lesion mask in MNI space
nb1000-Organ	Extracted lesioned voxels in MNI space
nT1nb1000-Chimeric	Chimeric image in MNI space
T1-Chimeric	Chimeric image in native T1 space
T1-Mask	Binary lesion mask in native T1 space
mT1-Chimeric	Midline aligned chimeric image in T1 space
mT1-Mask	Midline aligned binary mask in T1 space
rmT1-Mask	Reflected midline aligned binary mask in T1 space
rmT1-Organ	Extracted contralesional voxels in midlined aligned T1 space
mT1-Organ	Organ reflected to occupy the original lesion space
mT1-Enant	Midline aligned enantiomorphically corrected image

Figure 3.2 - Table of abbreviations for figure 3.1.

The various images are named with the volume space first and their role second. All recipient images are T1 scans, whilst donor images are diffusion weighted b0 and b1000 sequences.

Using the unified segmentation-normalisation routine the forward (native space into MNI) and inverse (MNI into native space) transformation parameters are obtained for a T1 recipient brain image. The inverse parameters are then applied to the lesioned DWI b1000 sequence and binary lesion mask that occupy MNI space thus transforming these two images into the native T1 space of the recipient.

With all images in native T1 space, the signal intensities of the voxels from the b1000 sequence within the focal lesion are extracted – the organ – and transplanted into their corresponding locations in the T1 image. To balance the signal intensities of the recipient and donor images, the recipient voxels were modulated by the ratio of the mean voxel signal intensity (excluding the lesioned voxels) of the recipient and donor brains (Brett et al., 2001).

The chimeric brain therefore consists of a background recipient T1 brain image with a transplanted lesion – organ – of the donor DWI b1000 scan. The

transplantation of the lesion from the donor into the recipient brain does not require an exact one to one correspondence with regards location, instead the creation of a chimeric brain is to facilitate the assessment of the spatial normalisation routine in the presence of a lesion. The assessment of the spatial normalisation methods can be performed on 2 identical brain images except for the presence of a lesion.

3.2.2.3 Midline alignment

The symmetry of the brain lies between the two hemispheres in the mid-sagittal plane. The brain image needs to be aligned in the midline of the image volume so that homologous regions in the two hemispheres may be matched in the subsequent steps. This is achieved by first reflecting the brain image in its original state in the sagittal midline of the image volume to create its mirror image, enantiomer. Next the transformation required to bring the original image into co-registration with its mid sagittal enantiomer is calculated using the default settings of the co-registration routine in SPM8 (appendix B). This manoeuvre is robust to the location of the lesion as it is a linear transform applied to the whole brain volume (rigid body transformation). The parameters necessary to align the images along the volume's sagittal midline is therefore half of the derived transformation within that plane (midline alignment parameters). The resultant brain image should have its mid sagittal plane in line with the volume's midline.

3.2.2.4 Enantiomorphically corrected brain creation

To create the enantiomorphic corrected images, the artificially lesioned chimeric image is first centred by aligning the mid-sagittal plane to the midline of the entire image with the transformation also applied to the associated binary lesion mask as described above (mT1-Chimeric, mT1-Mask).

The binary mask is then reflected in the mid sagittal plane (rmT1-Mask) to create an enantiomorphic copy and used to extract the healthy contralesional voxels (rmT1-Organ) from the chimeric image. The extracted signal intensities are then reflected back along the midline (mT1-Organ) and transplanted into the corresponding lesion space to create the enantiomorphically corrected image (mT1-Enant).

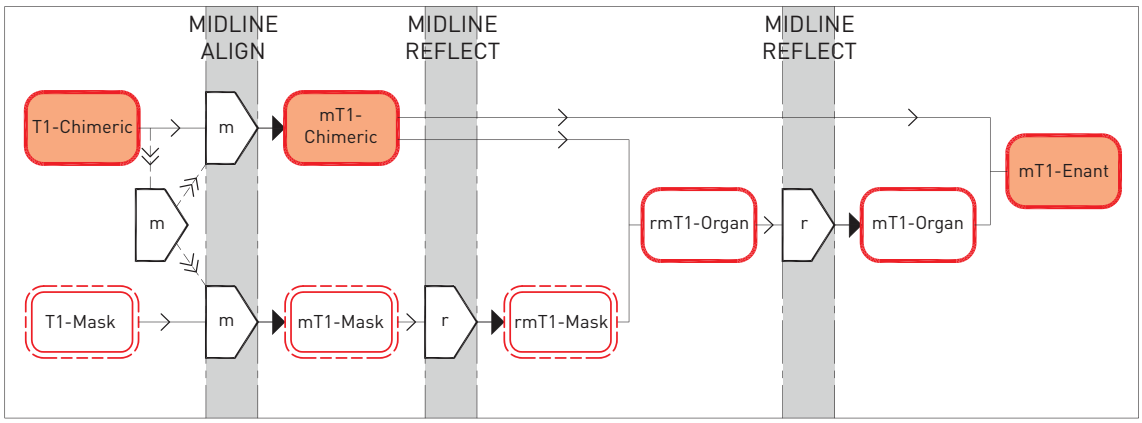


Figure 3.3 - Enantiomorphically corrected image creation.

An enantiomorphically corrected lesioned brain volume is created by aligning the lesioned brain volume and the associated binary mask in the mid-sagital plane. Next the midline aligned binary mask (mT1-Mask) is reflected in the mid-sagittal plane and used to extract the contralesional voxel signal intensities from the brain volume (rmT1-Organ). Finally the rmT1-Organ is reflected in the mid-sagittal plane and the injured voxels within mT1-Chimeric are replaced by the selected contralesion voxels to create the enantiomorphically corrected image (mT1-Enant).

To ensure a fair comparison of the various techniques all normalisation routines were assessed using midline aligned images rather than the original native space of the T1 image. Since the midline aligned image is the result of a rigid body transformation this should not affect the subsequent registration by the different routines. Unlike the other spatial normalisation techniques, the enantiomorphic method requires the creation of a pseudo healthy unlesioned brain by exploiting the symmetry in the mid-sagittal plane of the brain, and substitutes the lesioned voxels for their contralesional hemisphere counterparts.

3.2.3 Normalisation methods

The 2 main techniques assessed are the unified segmentation-normalisation routine and New segment, both part of the Statistical Parametric Mapping 8 (SPM8) software suite. From these 2 base approaches further modifications were applied – cost function masking (CFM), DARTEL and enantiomorphic normalisation – the complete list and their abbreviations used throughout this chapter are detailed below. Figure 3.5 and 3.6 illustrates the hierarchy of the methods and their associations. Unless specified the default SPM8 settings for the unified segmentation-normalisation, New segment and DARTEL algorithms were used. The specific parameter settings can be found in appendix B.

Method performed on unlesioned brains	Method performed on lesioned brains
Unified segmentation-normalisation with light regularisation **(ULP)**	Unified segmentation-normalisation with light regularisation **(ULP)** Unified segmentation-normalisation with light regularisation and cost function masking **(ULC)** Unified segmentation-normalisation with light regularisation and enantiomorphic correction **(ULE)**
Unified segmentation-normalisation with light regularisation and DARTEL **(ULPD)**	Unified segmentation-normalisation with light regularisation and DARTEL **(ULPD)** Unified segmentation-normalisation with light regularisation, cost function masking and DARTEL **(ULCD)** Unified segmentation-normalisation with light regularisation, enantiomorphic correction and DARTEL **(ULED)**
Unified segmentation-normalisation with medium regularisation **(UMP)**	Unified segmentation-normalisation with medium regularisation **(UMP)**
Unified segmentation-normalisation with medium regularisation and DARTEL **(UMPD)**	Unified segmentation-normalisation with medium regularisation and DARTEL **(UMPD)**
New segment with DARTEL **(NSPD)**	New segment with DARTEL **(NSPD)** New segment with enantiomorphic correction and DARTEL **(NSED)**
New segment with extra prior and DARTEL **(NWPD)**	New segment with extra prior and DARTEL **(NWPD)** New segment with extra prior, enantiomorphic correction and DARTEL **(NWED)**

Figure 3.4 - Table of abbreviations for the different normalisation methods.

All methods listed in the right hand column of the table were applied to the chimeric images, with the generated deformation fields compared with the deformation fields created by the corresponding method in the left hand column performed on the unlesioned T1 recipient brains.

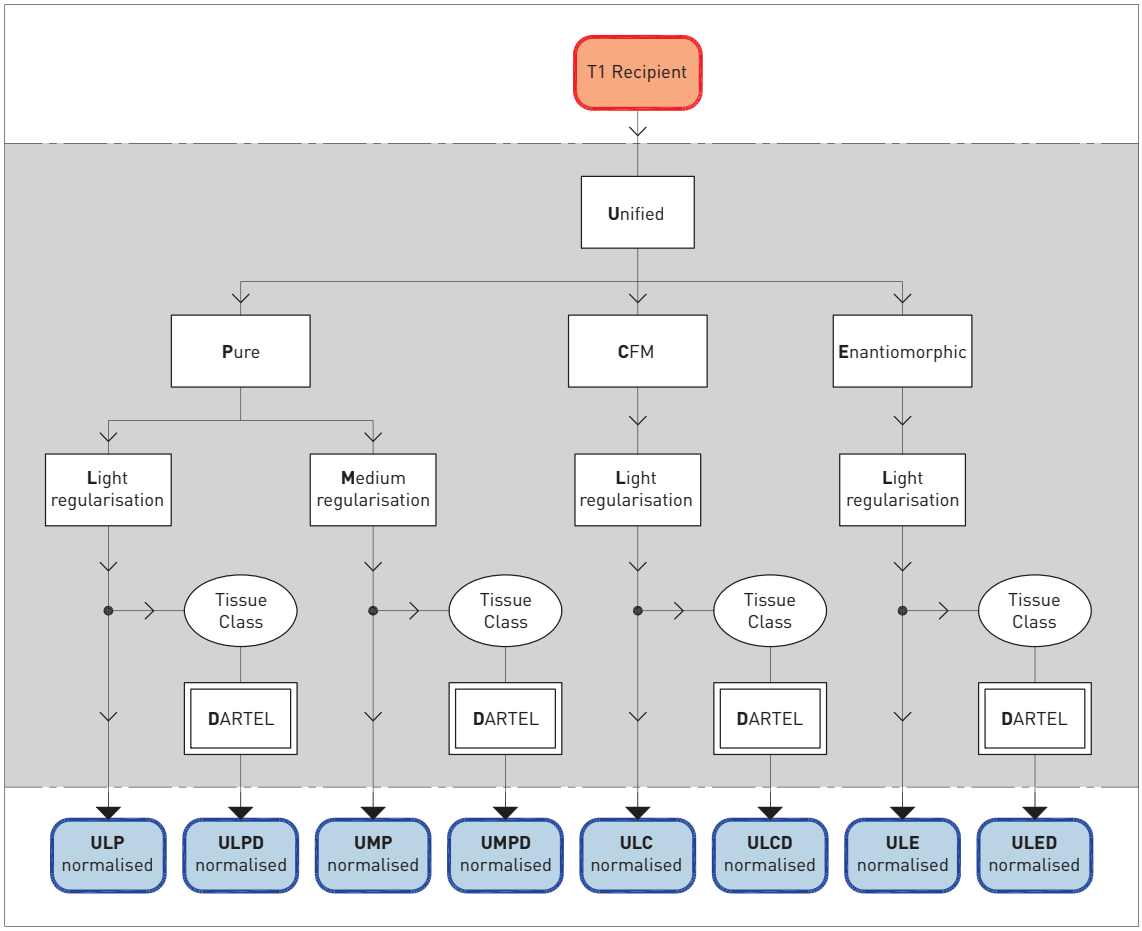


Figure 3.5 - Unified segmentation-normalisation routine base methods.

In addition to a pure unified segmentation-normalisation approach, either cost function masking (CFM) or the enantiomorphic modification was applied. All three subtypes were assessed with and without DARTEL.

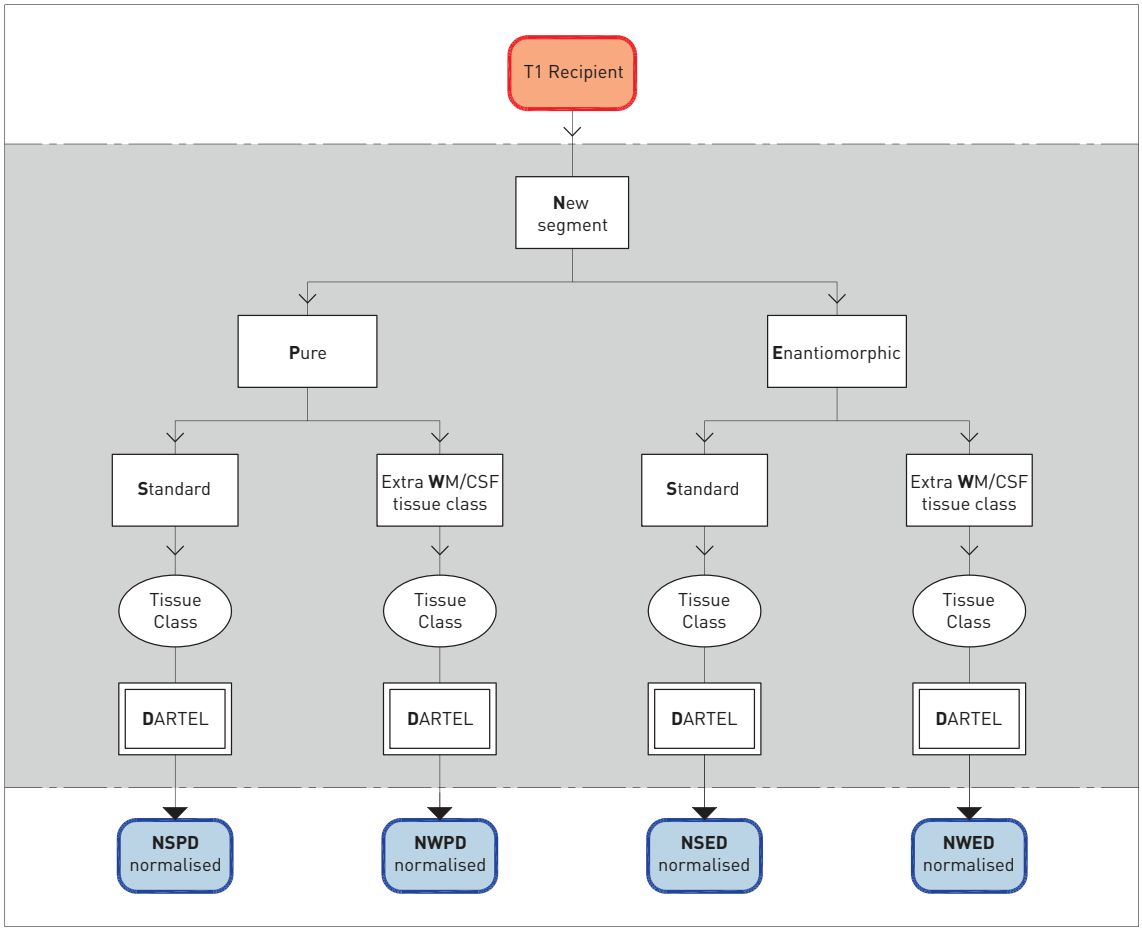


Figure 3.6 - New segment routine base methods.

Either a pure or enantiomorphic modification, both with DARTEL was applied. Both subtypes were assessed with and without the use of the additional tissue class prior created by taking the mean of the white matter and cerebrospinal fluid (CSF) priors found in SPM8.

3.2.3.1 Unified segmentation-normalisation

The simplest method in this sub group is the unified segmentation-normalisation technique, where either a healthy T1 or chimeric brain volume is supplied to the algorithm. Importantly, normalisation of the healthy brain in this manner will act as the comparison baseline against which the cost function masking (CFM) and enantiomorphic variations will be compared with, since the influence of either of these is irrelevant in the absence of a lesion.

In the pure (without any additional modification) unified segmentation-normalisation technique the brain images were first processed using the segment function in SPM8 with the bias regularisation set to either 0.001 (light) or 0.01 (medium). The segmented tissue classes were then passed to the normalisation function which generated a parameter file from which the final deformation fields were extracted (ULP, UMP).

A branch off the pure arm was the application of the DARTEL (Diffeomorphic Anatomical Registration Through Exponential Lie algebra) modification. The tissue classes that were extracted during the segmentation process in the pure arm were supplied to the DARTEL algorithm rather than to the normalisation function to create the DARTEL flow fields (ULPD, UMPD). From these flow fields the final deformation fields were extracted using the deformation function in SPM8.

In the cost function masking arm (CFM) the midline aligned chimeric images were first processed using the segment function in SPM8 where light bias regularisation was used to create the segmented tissue classes. Similar to the method used in Brett et al (Brett et al., 2001) the corresponding binary lesion mask was supplied to the segment function as the “masking image”, to ensure that the voxels contained within the lesion do not contribute to the estimation of the normalisation parameters. The tissue classes were then passed to the

normalisation function which generated a parameter file from which the final deformation fields were extracted (ULC).

A branch off the CFM arm was the application of the DARTEL modification, where the tissue classes created by the segment function were supplied to DARTEL rather than the normalisation function (ULCD). The resultant DARTEL flow fields were then used to extract the final deformation fields.

In the enantiomorphic arm the midline aligned enantiomorphic corrected images were first processed using the segment function in SPM8 where light bias regularisation was used to create the segmented tissue classes (ULE). The tissue classes were then passed to the normalisation function which generated a parameter file from which the final deformation fields were extracted.

The segmented tissue classes were supplied to the DARTEL modification in a branch off the enantiomorphic method. The resultant DARTEL flow fields were then used to create the final deformation fields (ULED).

3.2.3.2 New segment

The New segment function (Ashburner and Friston, 2005) was also used to create the segmented tissue classes rather than the standard segment function in SPM8 and passed to the DARTEL modification. In all cases the default settings for New segment found in SPM8 were used (appendix B).

In the pure arm (without additional priors or enantiomorphic corrected images) the unlesioned T1 images and chimeric images were processed with the New segment function. The segmented tissue classes created specifically for DARTEL (grey matter, white matter and CSF) were then processed using the DARTEL modification (NSPD). The resultant DARTEL flow fields were then used

to generate the final deformation field via the deformations function found in SPM8.

An additional prior was used with the New segment routine as in Seghier et al (2008) (Extra white matter / cerebrospinal fluid (WM/CSF) tissue class arm). The extra tissue class was derived from the mean of the white matter and CSF priors found in the SPM8 suite (NWPD). The resultant segmented classes were then passed to DARTEL to determine the final deformation fields.

The New segment enantiomorphic arm, utilised the enantiomorphic corrected images and were subsequently processed using the New segment function with the standard priors (NSED). Once again the segmented tissue classes were passed to DARTEL to create the DARTEL flow fields, which were then used to derive the final deformation fields via the deformations function.

The final variation was the use of the extra WM/CSF prior with New segment, in combination with the enantiomorphic corrected images and DARTEL (NWED). The resultant DARTEL flow fields were used to create the final deformation fields via the deformations function.

3.2.4 Evaluation

The spatial normalisation process can be specified as a field of three dimensional vectors at every voxel describing where the voxel ends up from the source image to the template. This is a deformation field. In the absence of a ground truth, the deformation field derived from the chimeric brain with the focal lesion is compared with that derived from the original T1 brain without the lesion. In the perfect situation the normalisation method will be unaffected by the presence of the lesion, therefore the deformation fields of the 2 approaches should be identical. We can quantify this displacement by calculating the

difference between the Euclidean distances of the vectors at each voxel. A summary measure for each image describing the discrepancy between the 2 scenarios can be quantified as the voxel-wise root mean squared difference of the images.

$$D_n = \sqrt{\left(x_n^{chimeric} - x_n^{unlesioned}\right)^2 + \left(y_n^{chimeric} - y_n^{unlesioned}\right)^2 + \left(z_n^{chimeric} - z_n^{unlesioned}\right)^2}$$

$$RMSD = \sqrt{\frac{\sum_{n=1}^{T} D^2}{T}}$$

The larger the root mean squared difference the greater the effect of the lesion on perturbing the normalisation process.

There are a total of 12 different spatial normalisation techniques assessed in the above experiment. The healthy unlesioned T1 images were processed along the pure arms in the unified segmentation-normalisation with and without the DARTEL modification and New segment base methods with the DARTEL modification. The 4 variations with the unlesioned images generated the deformation fields against which all 12 variations were compared against.

There is a linear relationship between the base 10 logarithm of the root mean squared difference ($\log_{10}$(RMSD)), and the base 10 logarithm of lesion volume ($\log_{10}$(lesion volume)) (Nachev et al. 2008). Therefore to assess whether there was an interaction between lesion volume and the spatial normalisation technique an analysis of covariance (ANCOVA) was performed. Similar to the method used in Nachev et al (2008), the response variable was $\log_{10}$(RMSD), for each collection of 11 chimeric images derived from a single b1000 lesion image and the 11 recipient MRI T1 unlesioned images. The covariate was the $\log_{10}$(lesion volume). The analysis was performed in MATLAB using the built-in

aoctool function and setting the function to generate separate lines rather than parallel lines.

During the normalisation process the algorithm will endeavour to try and minimize the differences between the test image and the standard template. Generally this will involve varying amounts of warping at different locations within the volume. Moreover, since the algorithm is trying to maximize the correspondence – of normal tissue – between the two images, there is a concern that it will penalize focal lesions disproportionately and tend to shrink the total volume of the lesioned tissue. To assess this feature, the binary mask is transformed to MNI space by the normalisation parameters derived from the unlesioned brain (T1 recipient brain) and the associated chimeric brain. The ratio of these two volumes that inhabit MNI space (binary mask volume$_{\text{unlesioned brain}}$ / binary mask volume$_{\text{chimeric brain}}$) is subsequently calculated, with values greater than one representing a reduction and less than one representing an expansion of the binary mask volume compared with what would be expected. This ratio is later referred to as the *volume change ratio*.

The ANCOVA analyses allow different methods to be compared and can identify whether there is a significant difference between methods; an effect of lesion volume; and if there is an interaction between lesion volume and method.

For each group of 11 chimeric brains that possess the same transplanted lesion, a 2 sample *t*-test was performed to compare different methods. The methods can be compared with respects to both RMSD and volume change ratio. This was performed in the situation where the two ANCOVA lines would intersect to help identify a lesion volume range in which the methods may differ significantly in performance.

With respects to the volume change ratio, it is the absolute deviation from 1 that is of interest. Consequently in the 2 sample *t*-test the modulus of the deviation from 1 was calculated prior to performing the 2 sample *t*-test.

3.3 Results

There are a total of 12 different spatial normalisation techniques under investigation. Since the analyses rely on the use of artificially lesioned chimeric brains, their performances can either be performed in a recipient wise, or a donor wise manner.

The volume change ratio describes whether the lesion undergoes expansion (less than 1) or shrinkage (greater than 1) after the normalisation process. Only the methods that used the enantiomorphic modification (ULE, ULED, NSED and NWED) and the New segment with an additional prior and DARTEL (NWPD) had a mean percentage volume change less than 1. All other techniques would tend to shrink the lesion volume.

The techniques which modified the resultant lesion volume the least were ULE, ULC and ULED with mean percentage volume changes of 0.11%, 0.13% and 0.13% respectively.

3.3.1 Assessment partitioned according to the background (recipient) brain scan

Across the 3 base methods (unified segmentation-normalisation, unified segmentation-normalisation with DARTEL and New Segment with DARTEL), there is a range of 0.1385 – 2.504mm for the mean RMSD across all methods. Figures 3.7, 3.8 and 3.9 show the 11 mean RMSD and standard errors of the 77 different donor lesions transplanted into each recipient brain for the unified segmentation-normalisation, unified segmentation-normalisation with DARTEL and New segment with DARTEL respectively.

All methods have a range below 1mm, with the 8 unified segmentation-normalisation methods having ranges below 0.4mm. Both the unified

segmentation-normalisation with light regularisation, enantiomorphic modification with and without DARTEL (ULE, ULED) exhibits the smallest ranges of 0.1811mm and 0.1851 respectively.

The smallest inter-quartile range was achieved with the unified segmentation-normalisation with light regularisation without DARTEL (0.0401mm), while both ULE and ULED had an inter-quartile range of 0.0651mm and 0.0586mm.

Within the New segment with DARTEL base method, NSED had the smallest range and inter-quartile range of 0.3972mm and 0.0885mm, amongst the 4 variations.

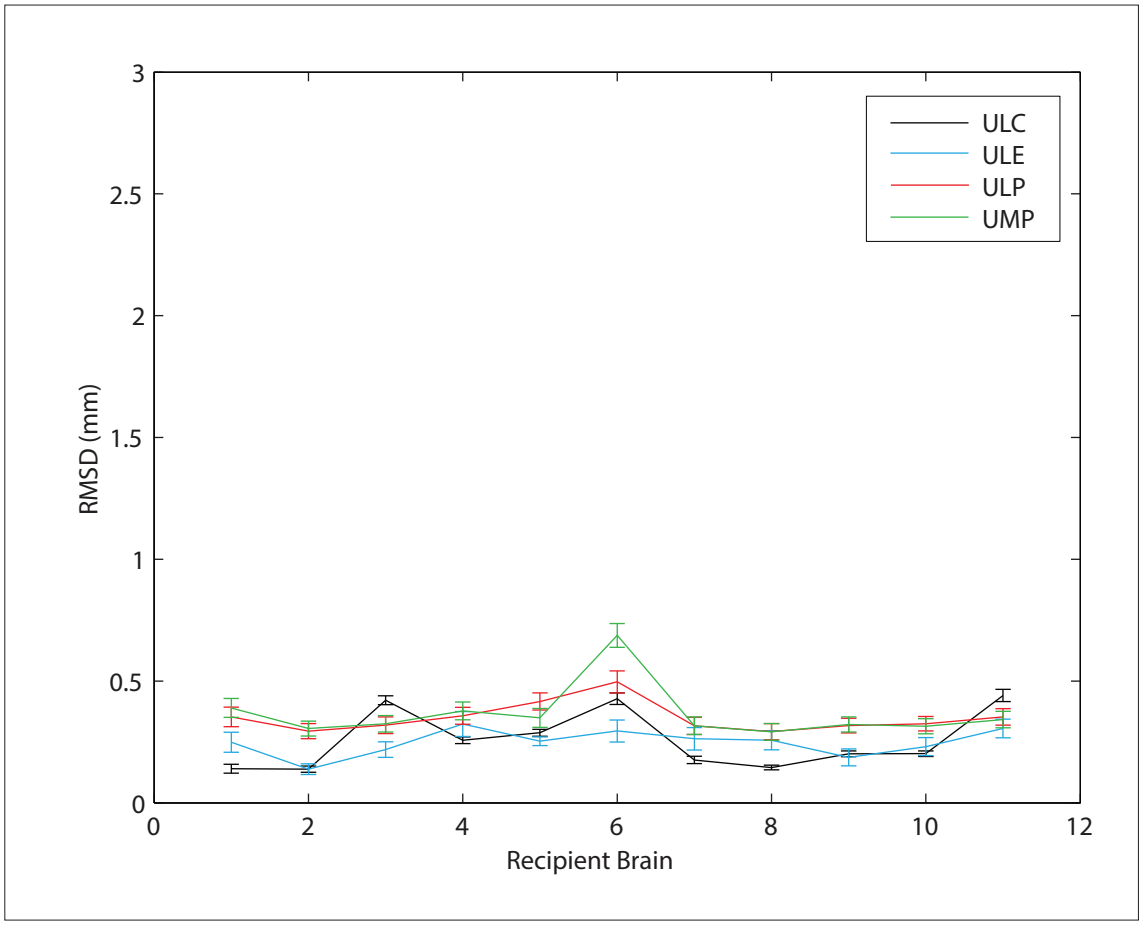


Figure 3.7 - Unified segmentation-normalisation routines (recipient): mean root mean squared differences.

Mean root mean squared difference (mm) for different spatial normalisation techniques using the base unified segmentation-normalisation routine for each set of 77 chimeric images with the same background T1 image. Error bars represent on standard error. The RMSD is calculated by first subtracting the deformation fields of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted.

Black: Light regularisation with cost function masking (ULC)

Blue: Light regularisation with enantiomorphic normalisation (ULE)

Red: Light regularisation (ULP)

Green:Medium regularisation (UMP)

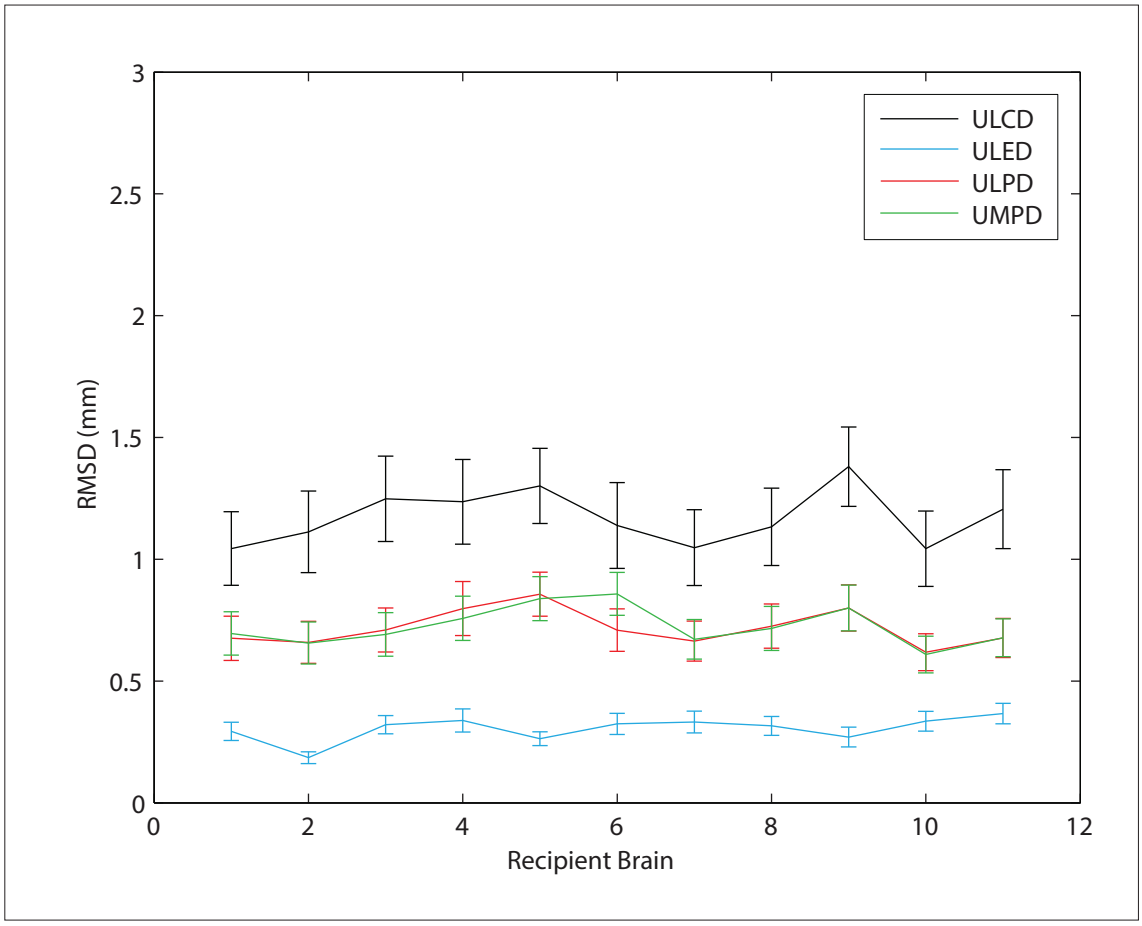


Figure 3.8 - Unified segmentation-normalisation and DARTEL routines (recipient): mean root mean squared differences.

Mean root mean squared difference (mm) for different spatial normalisation techniques using the base unified segmentation-normalisation and DARTEL routine for each set of 77 chimeric images with the same background T1 image. Error bars represent on standard error. The RMSD is calculated by first subtracting the deformation fields of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted.

Black: Light regularisation with cost function masking (ULCD)

Blue: Light regularisation with enantiomorphic normalisation (ULED)

Red: Light regularisation (ULPD)

Green:Medium regularisation (UMPD)

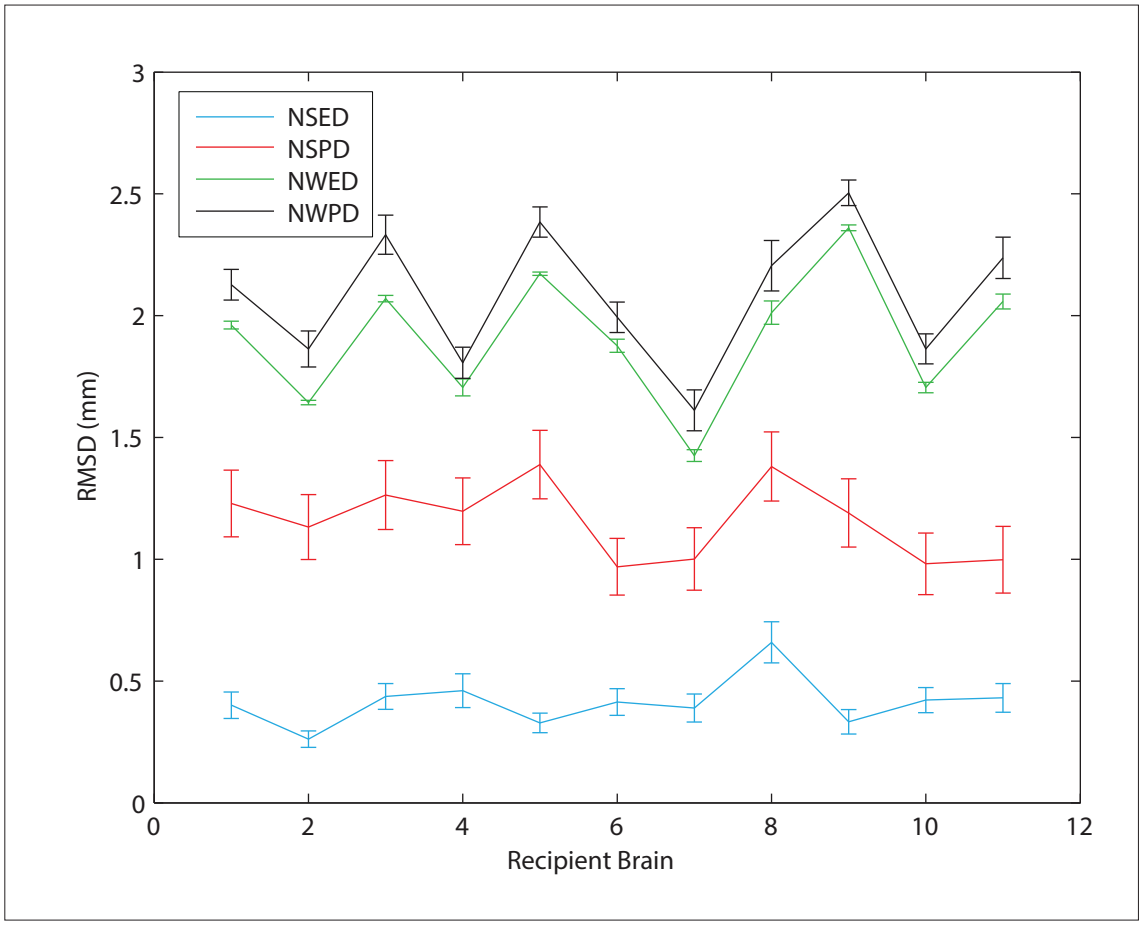


Figure 3.9 - New segment and DARTEL routines (recipient): mean root mean squared differences.

Mean root mean squared difference (mm) for different spatial normalisation techniques using the base New segment and DARTEL routine for each set of 77 chimeric images with the same background T1 image. Error bars represent on standard error. The RMSD is calculated by first subtracting the deformation fields of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted.

Blue: Enantiomorphic normalisation (NSED)

Red: No further modifications (NSPD)

Green: Extra WM/CSF prior and enantiomorphic normalisation (NWED)

Black: Extra white matter/CSF (WM/CSF) prior (NWPD)

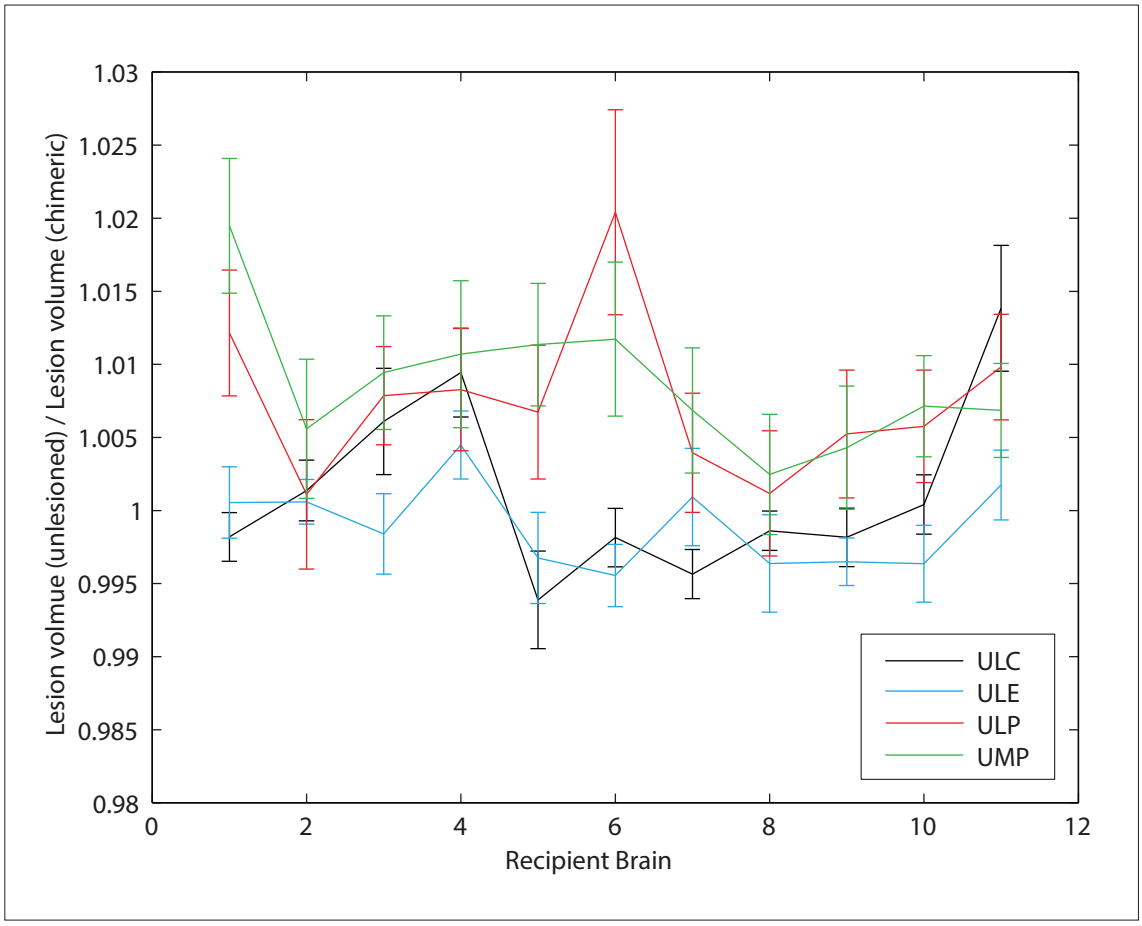


Figure 3.10 - Unified segmentation-normalisation routines (recipient): mean volume change ratios.

Mean volume change ratio for different spatial normalisation techniques using the base unified segmentation-normalisation routine for each set of 77 chimeric with the same background T1 image. Error bars represent on standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Black: Light regularisation with cost function masking (ULC)

Blue: Light regularisation with enantiomorphic normalisation (ULE)

Red: Light regularisation (ULP)

Green:Medium regularisation (UMP)

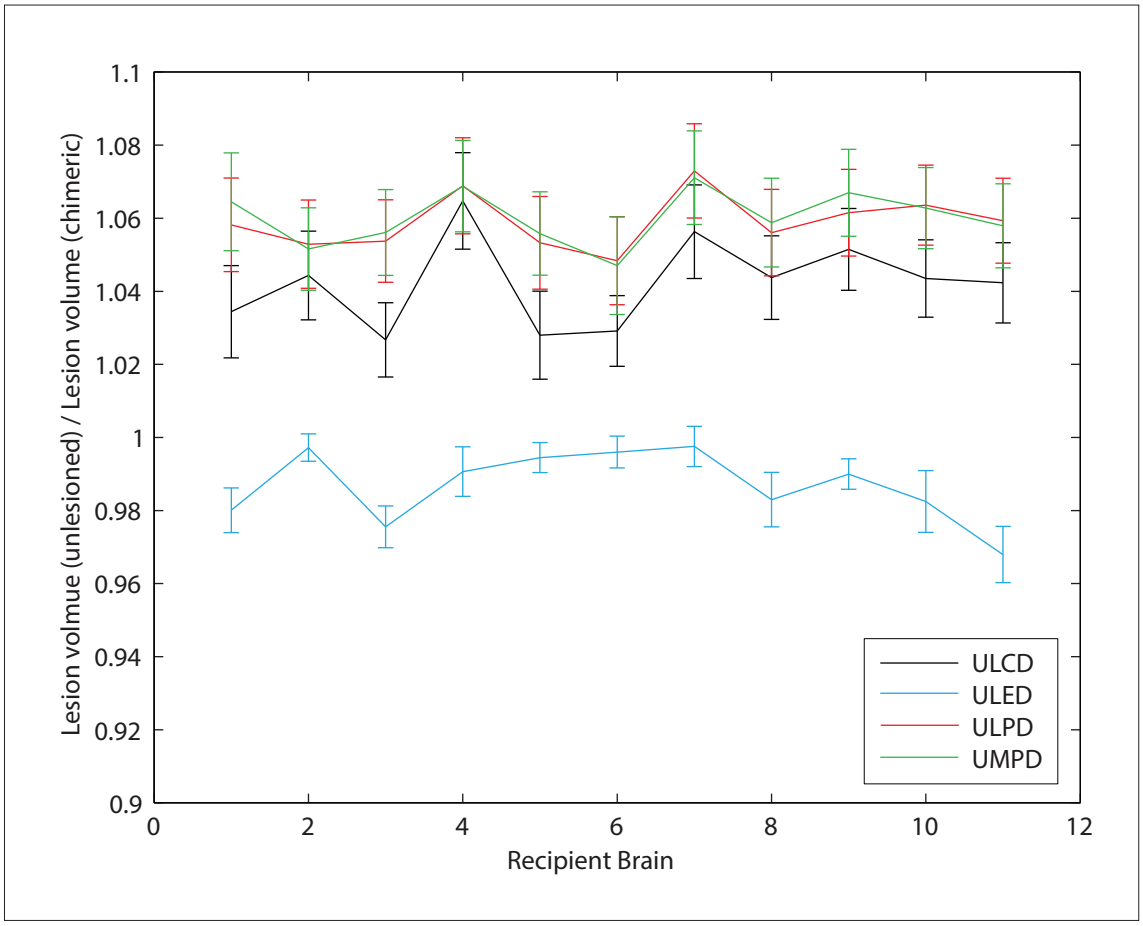


Figure 3.11 - Unified segmentation-normalisation and DARTEL routines (recipient): mean volume change ratios.

Mean volume change ratio for different spatial normalisation techniques using the base unified segmentation-normalisation and DARTEL routine for each set of 77 chimeric images with the same background T1 image. Error bars represent on standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Black: Light regularisation with cost function masking (ULCD)

Blue: Light regularisation with enantiomorphic normalisation (ULED)

Red: Light regularisation (ULPD)

Green:Medium regularisation (UMPD)

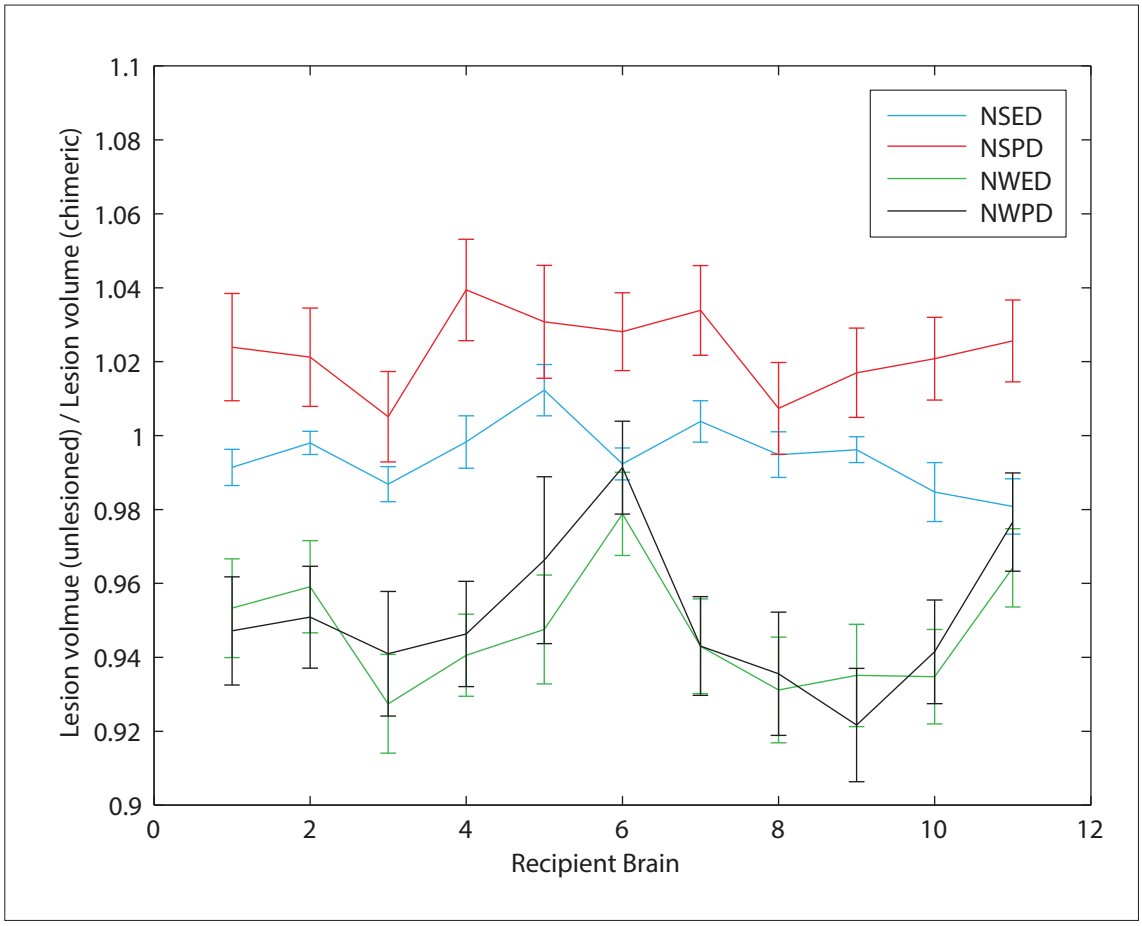


Figure 3.12 - New segment and DARTEL routines (recipient): mean volume change ratios.

Mean volume change ratio for different spatial normalisation techniques using the base New segment and DARTEL routine for each set of 77 chimeric images with the same background T1 image. Error bars represent on standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Blue: Enantiomorphic normalisation (NSED)

Red: No further modifications (NSPD)

Green: Extra WM/CSF prior and enantiomorphic normalisation (NWED)

Black: Extra white matter/CSF (WM/CSF) prior (NWPD)

The volume change ratio represents the amount the spatial normalisation technique either enlarges (less than 1) or shrinks (greater than 1) the original lesion volume. The lesion mask in native T1 space can be transformed into standard MNI space by using either the normalisation parameters from the unlesioned recipient brain or from the complementary chimeric brain. The volume change ratio is therefore the volume ratio of unlesioned : chimeric, and provides an alternative measure of how much the normalisation technique is perturbed by the presence of a lesion.

The recipient brain has little effect on this measure, irrespective of the technique, with all techniques having an inter quartile range of less than 2.5%.

The unified segmentation-normalisation with enantiomorphic normalisation (ULE) has the smallest range at 0.89% (figure 3.10), while UMP, ULP and ULC have a range of 1.7%, 1.9% and 2.0% respectively.

3.3.2 Assessment partitioned according to the lesion (donor) brain scan

The 77 different lesion volumes are arranged in ascending order along the *x* axis on a base 10 logarithmic scale, $\log_{10}$(lesion volume), and the resultant RMSD plotted on a base 10 logarithmic, $\log_{10}$(RMSD), scale on the *y* axis. Across all 12 techniques there is a linear correlation between lesion volume and the RMSD, with larger lesions being associated with larger RMSD scores.

The root mean squared difference is calculated by comparing the deformation fields generated by applying the normalisation method to the chimeric image and its healthy unlesioned recipient image. The Euclidean distance of the difference in each plane is calculated and the sum of the squared Euclidean distances across all voxels is derived. Therefore the summary statistic for the entire brain is therefore the mean of the summed squared Euclidean distances

all square rooted. The chimeric brains are grouped according to the common lesion brain with the mean and standard error plotted on a $\log_{10}$(RMSD) by $\log_{10}$(lesion volume) graph. All methods display a significant correlation between the RMSD and lesion volume except for the New segment techniques that utilise an extra white matter and CSF prior.

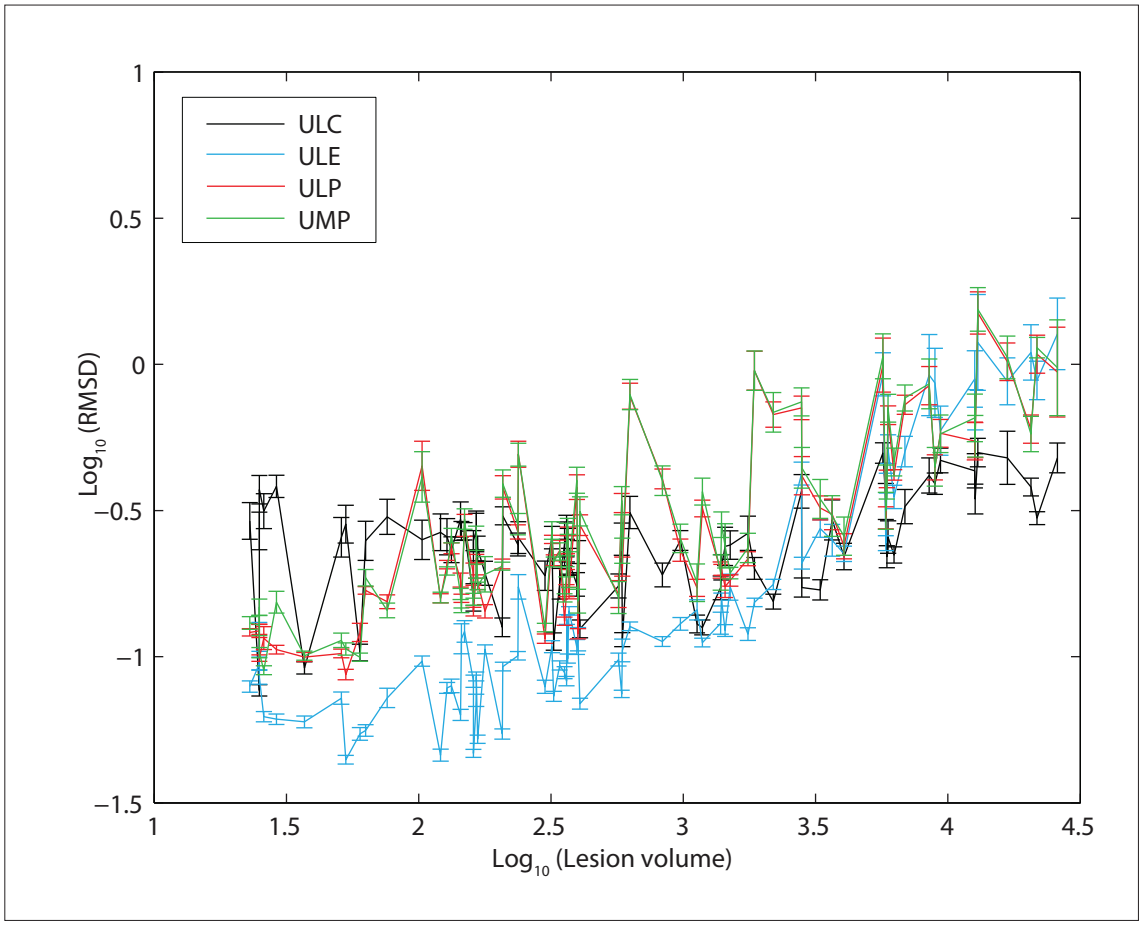


Figure 3.13 - Unified segmentation-normalisation routines: Log_{10}(RMSD) vs Log_{10}(lesion volume).

The RMSD is calculated by first subtracting the deformation fields (mm) of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted. The base 10 logarithm of the mean RMSD for each set of 11 chimeric images with the same donor b1000 lesion is plotted against the base 10 logarithm of the donor lesion volume (voxels). Error bars represent one standard error. As the size of the lesion volume increases so does the RMSD for all methods that use the base unified segmentation-normalisation routine.

Black: Light regularisation with cost function masking (ULC)

Blue: Light regularisation with enantiomorphic normalisation (ULE)

Red: Light regularisation (ULP)

Green:Medium regularisation (UMP)

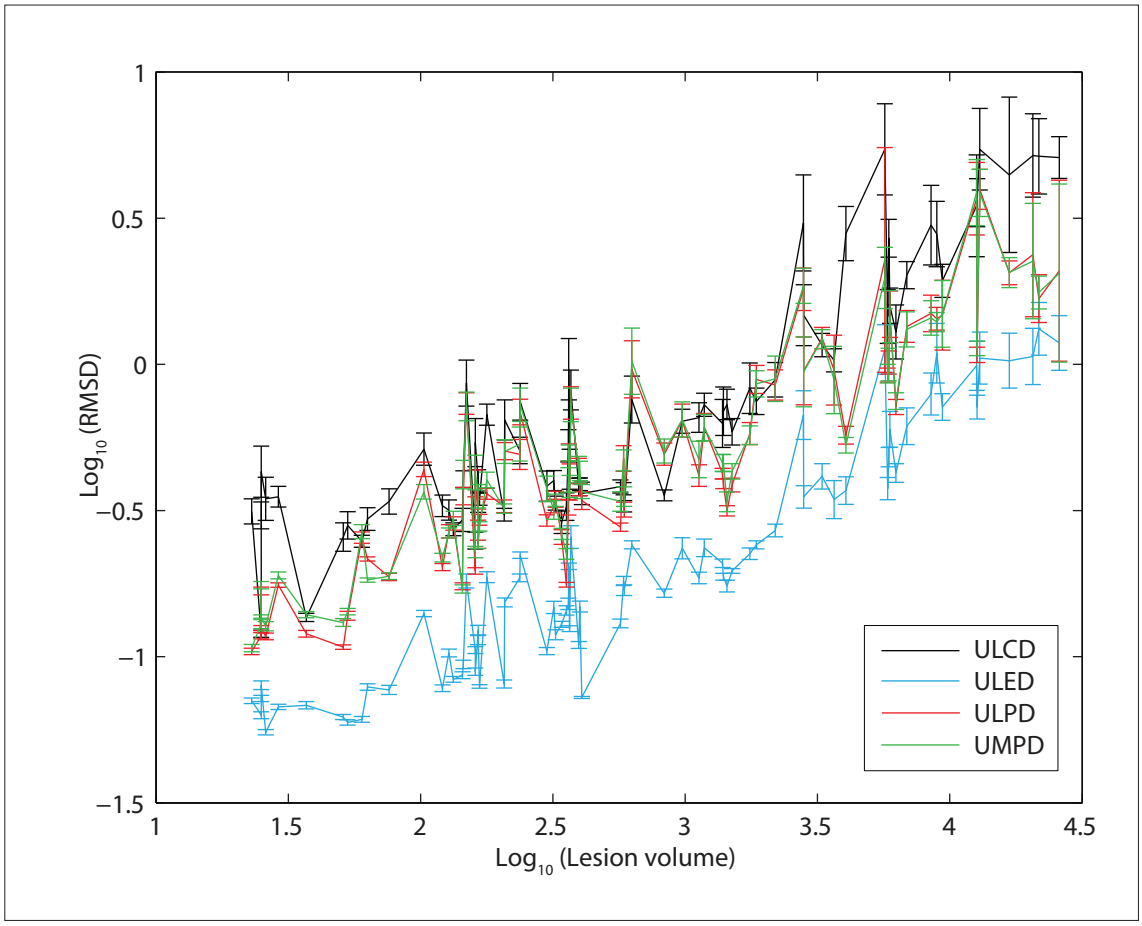


Figure 3.14 - Unified segmentation-normalisation and DARTEL routines: Log_{10}(RMSD) vs Log_{10}(lesion volume).

The RMSD is calculated by first subtracting the deformation fields (mm) of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted. The base 10 logarithm of the mean RMSD for each set of 11 chimeric images with the same donor b1000 lesion is plotted against the base 10 logarithm of the donor lesion volume (voxels). Error bars represent one standard error. As the size of the lesion volume increases so does the RMSD for all methods that use the base unified segmentation-normalisation and DARTEL routine.

Black: Light regularisation with cost function masking (ULCD)

Blue: Light regularisation with enantiomorphic normalisation (ULED)

Red: Light regularisation (ULPD)

Green:Medium regularisation (UMPD)

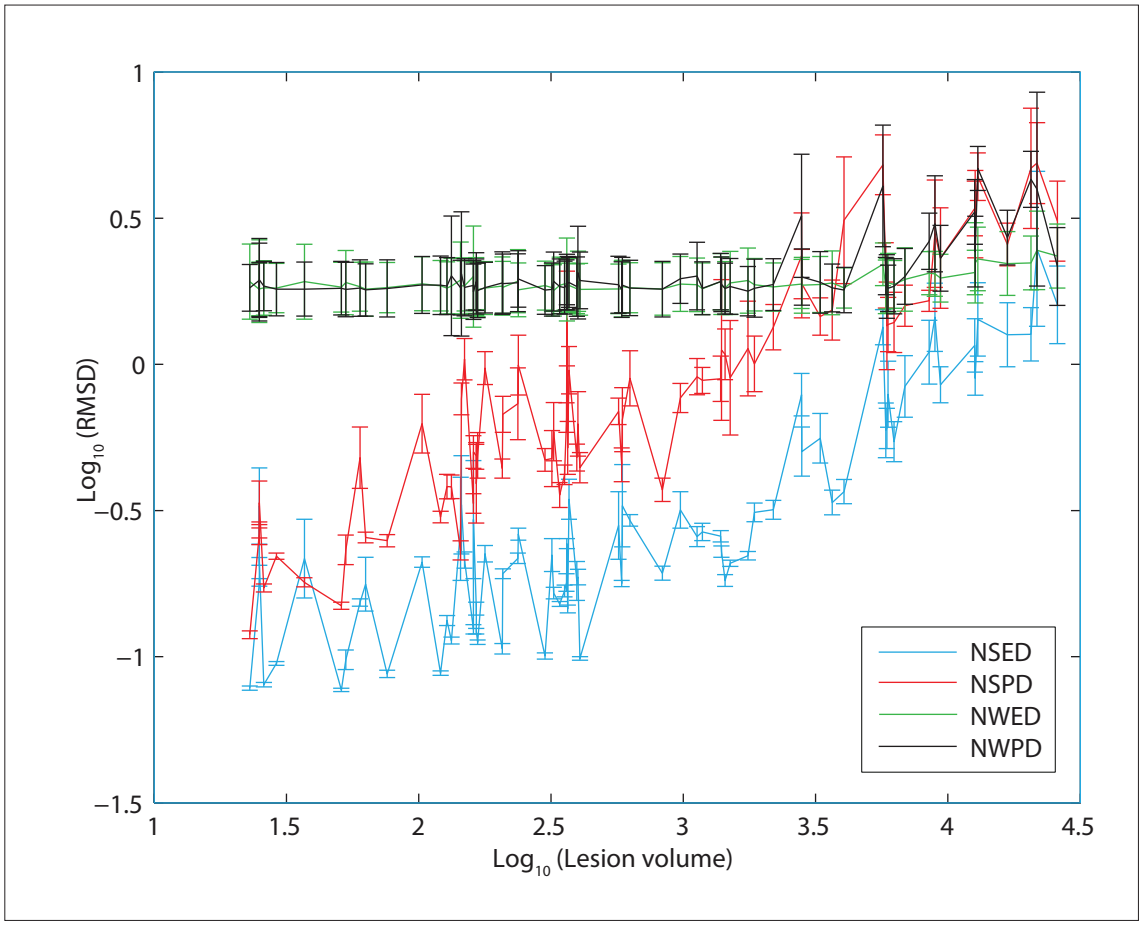


Figure 3.15 - New segment and DARTEL routines:
Log_{10}(RMSD) vs Log_{10}(lesion volume).

The RMSD is calculated by first subtracting the deformation fields (mm) of the chimeric image from its healthy unlesioned T1 recipient brain. The difference at each voxel is then squared, with the sum of the squared difference finally square rooted. The base 10 logarithm of the mean RMSD for each set of 11 chimeric images with the same donor b1000 lesion is plotted against the base 10 logarithm of the donor lesion volume (voxels). Error bars represent one standard error. The base method is New segment and DARTEL routine. There is no effect of lesion volume on the RMSD for methods that utilise the additional WM/CSF prior.

Blue: Enantiomorphic normalisation (NSED)

Red: No further modifications (NSPD)

Green: Extra WM/CSF prior and enantiomorphic normalisation (NWED)

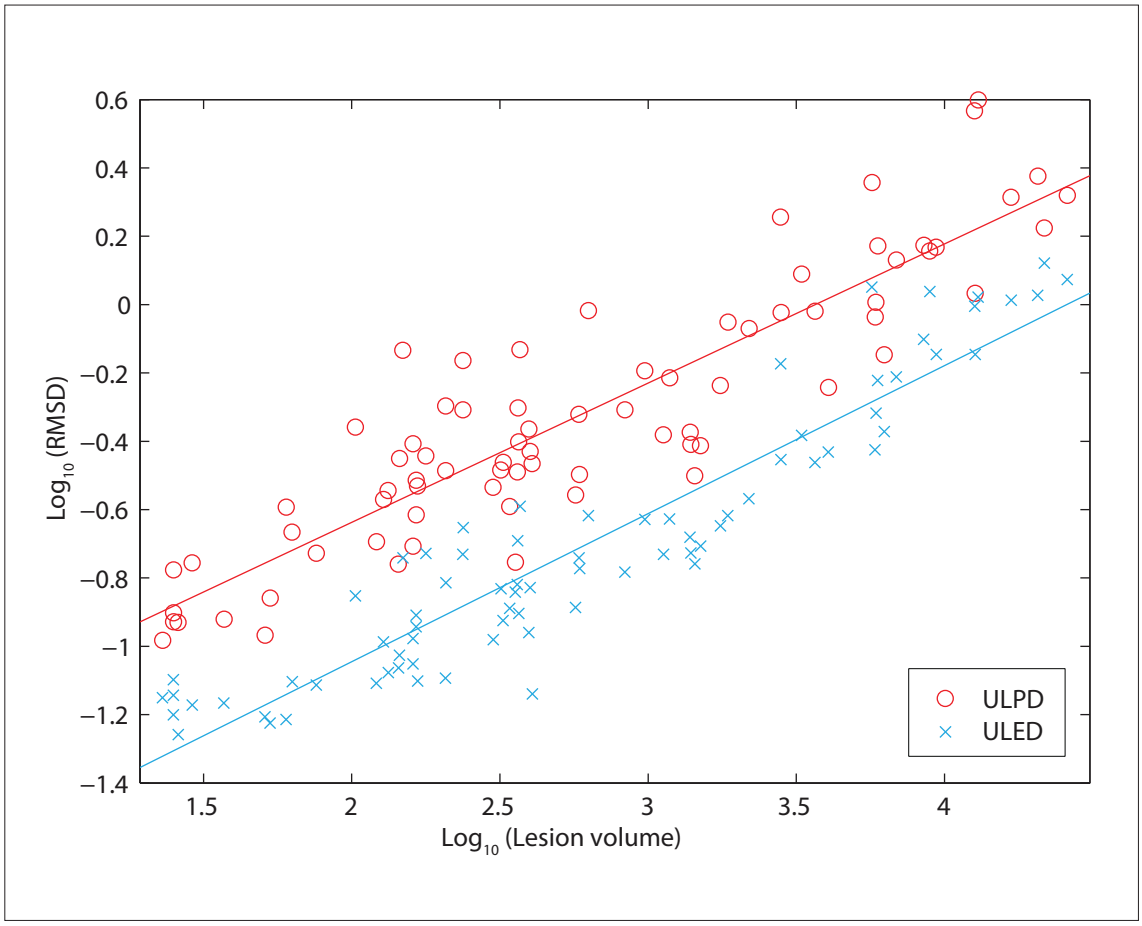
Black: Extra white matter/CSF (WM/CSF) prior (NWPD)

The methods with the smallest overall mean RMSDs are ULC, ULE, ULED and NSED, at 0.2580mm, 0.2473mm, 0.3040mm and 0.4123mm.

In all the following ANCOVA analyses the different normalisation techniques were treated as the co-variable.

Figure 3.16 compares the unified segmentation-normalisation with DARTEL in the presence and absence of the enantiomorphic modification (ULPD, ULED), and figure 3.17 compares the New segment with DARTEL in the presence and absence of the enantiomorphic normalisation (NSPD, NSED). In both cases, whilst in the presence of DARTEL, the addition of the enantiomorphic modification reduces the RMSD for the technique across the entire lesion volume range, without causing an interaction between method and lesion volume.

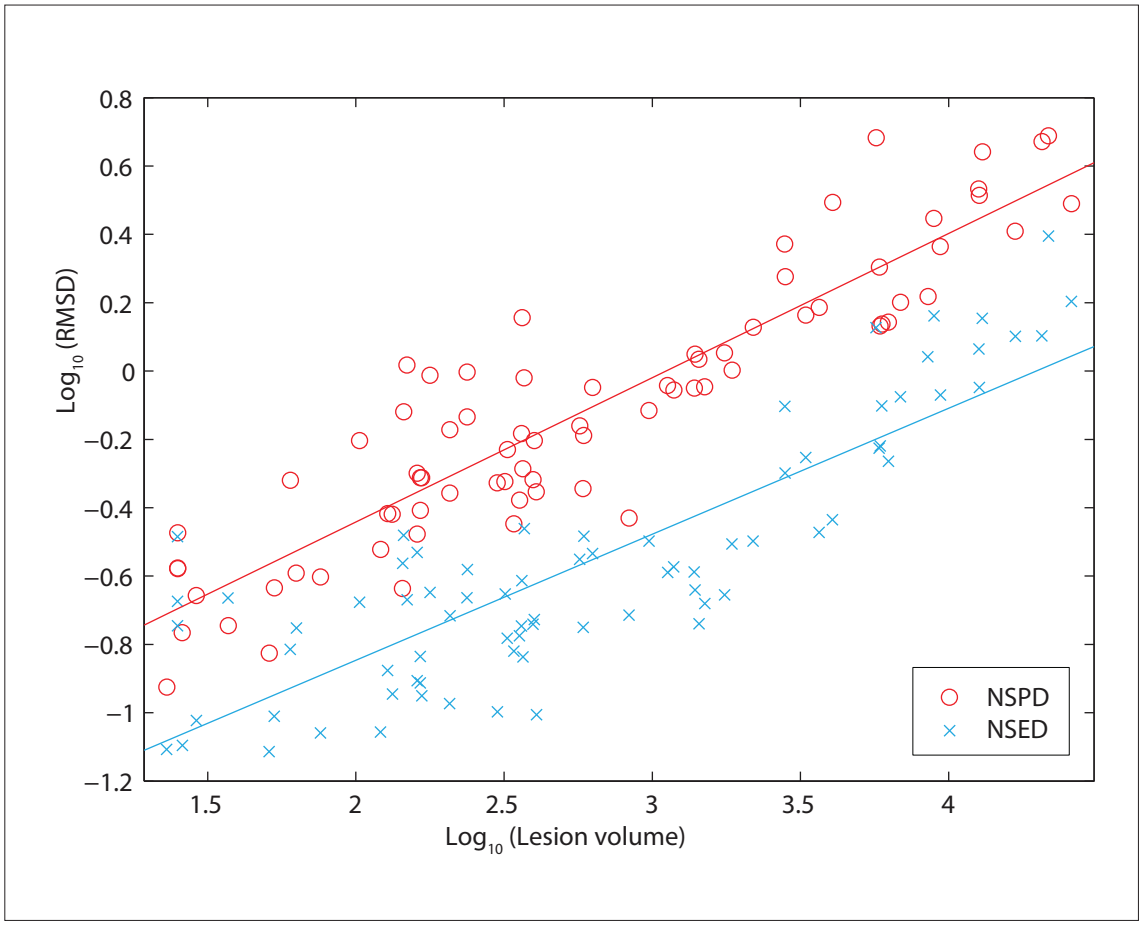
If it is assumed that the lines are parallel in the 2 cases, as there is no significant interaction between method and lesion volume across the dataset used, the separation between ULPD and ULED is $10^{0.3876}$ (2.441mm), whilst the separation between NSPD and NSED is $10^{0.4464}$ (2.795mm).



ULPD vs. ULED	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	5.78	5.78	252.60	0.00E+00
Lesion volume	1	18.90	18.90	825.52	0.00E+00
Method x Lesion volume	1	0.02	0.02	0.78	3.80E-01
Error	150	3.43	0.02		

Figure 3.16 - ANCOVA comparing ULPD and ULED.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation and DARTEL method in the absence (red) and presence (blue) of the enantiomorphic modification. The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is no significant interaction between method and lesion volume.

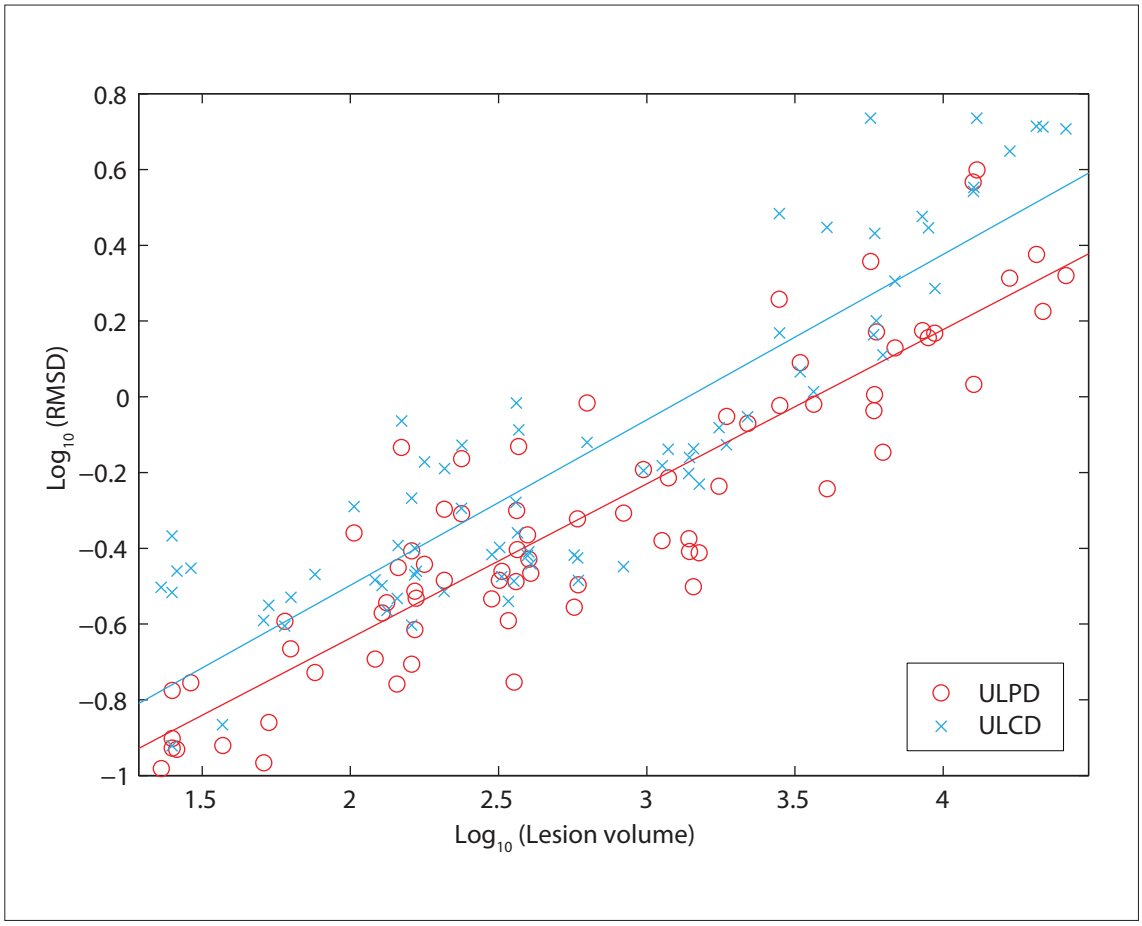


NSPD vs. NSED	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	7.67	7.67	240.34	0.00E+00
Lesion volume	1	16.73	16.73	524.10	0.00E+00
Method x Lesion volume	1	0.08	0.08	2.39	1.24E-01
Error	150	4.79	0.03		

Figure 3.17 - ANCOVA comparing NSPD and NSED.

ANCOVA analysis comparing the New segment and DARTEL method in the absence (red) and presence (blue) of the enantiomorphic modification. The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is no significant interaction between method and lesion volume.

In contrast, the addition of cost function masking to the unified segmentation-normalisation with DARTEL technique (ULCD) results in an increase in the RMSD across the lesion range represented as an upward shift from the ULPD line in figure 3.18. The two methods are significantly different, but there is no significant interaction between lesion volume and method (p=0.39).



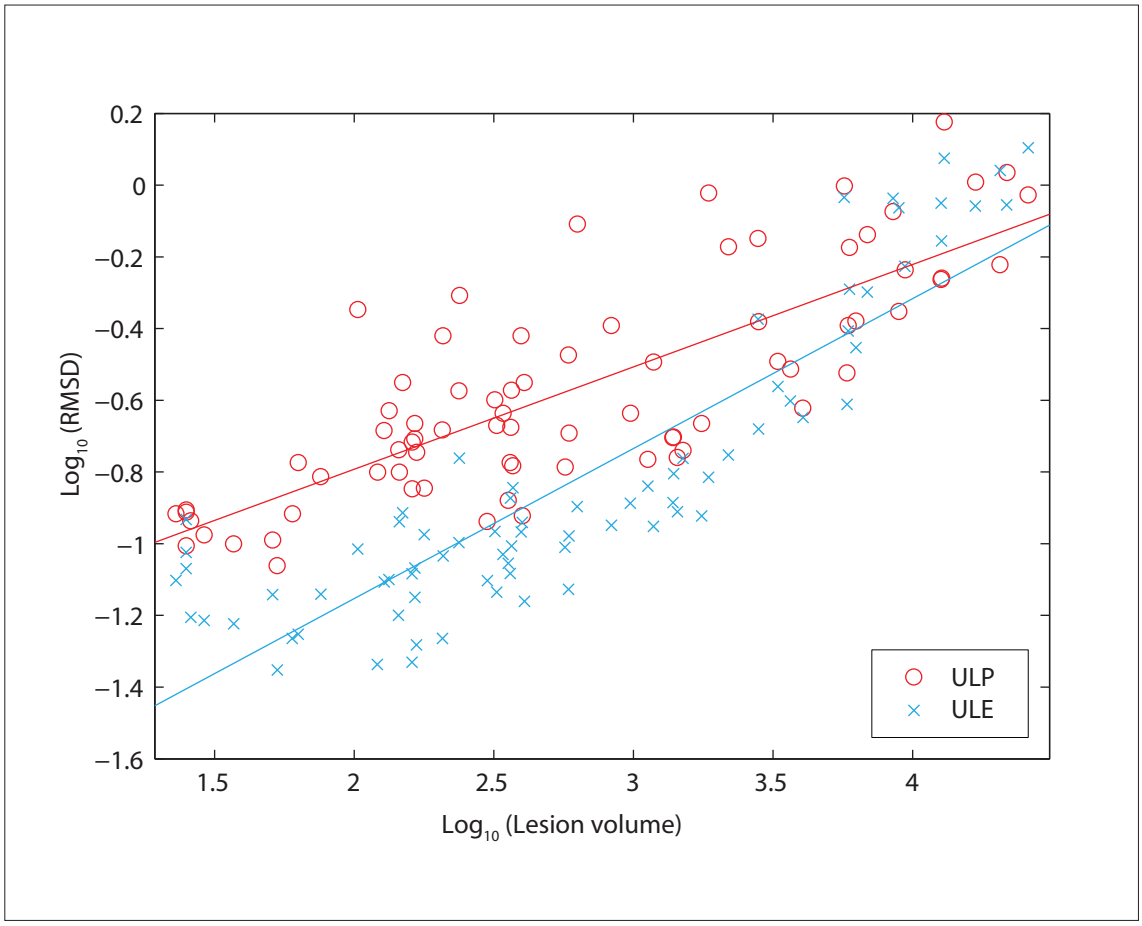
ULPD vs. ULCD	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	1.02	1.02	31.47	0.00E+00
Lesion volume	1	19.07	19.07	589.02	0.00E+00
Method x Lesion volume	1	0.02	0.02	0.72	3.98E-01
Error	150	4.86	0.03		

Figure 3.18 - ANCOVA comparing ULPD and ULCD.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation and DARTEL with (red) and without (blue) cost function masking. The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is no significant interaction between method and lesion volume. The two lines are almost parallel with no significant interaction between lesion volume and method (p=0.39).

In the absence of DARTEL, the effect of the enantiomorphic modification is different from the effect of cost function masking, when used in combination with the unified segmentation-normalisation base method.

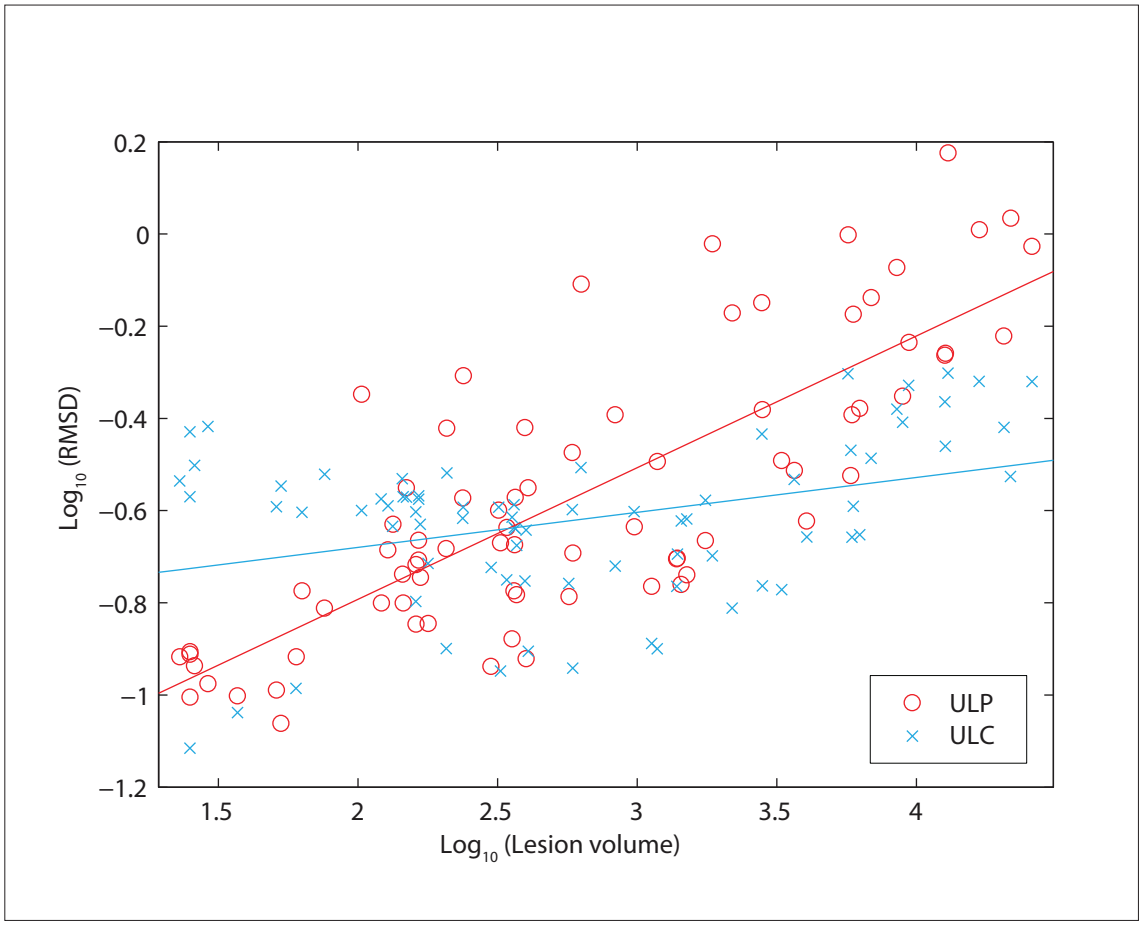
The ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation (ULP) with the unified segmentation-normalisation with light regularisation and enantiomorphic normalisation (ULE) reveals a reduction in RMSD particularly for small to medium sized lesions (figure 3.19). The effect of method alone and the interaction between lesion volume and method both reach significance (p=0.00, p=0.0003 respectively). The 2 lines intersect at approximately $10^{4.5}$ (30,000 voxels).



ULP vs. ULE	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	2.53	2.53	74.00	0.00E+00
Lesion volume	1	13.24	13.24	386.88	0.00E+00
Method x Lesion volume	1	0.47	0.47	13.77	3.00E-04
Error	150	5.1341	0.0342		

Figure 3.19 - ANCOVA comparing ULP and ULE.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation (red) against the unified segmentation-normalisation with light regularisation and enantiomorphic normalisation technique (blue). The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is a significant effect of method, lesion volume and interaction between the two. The two lines intersect just above $10^{4.5}$ (30,000 voxels).



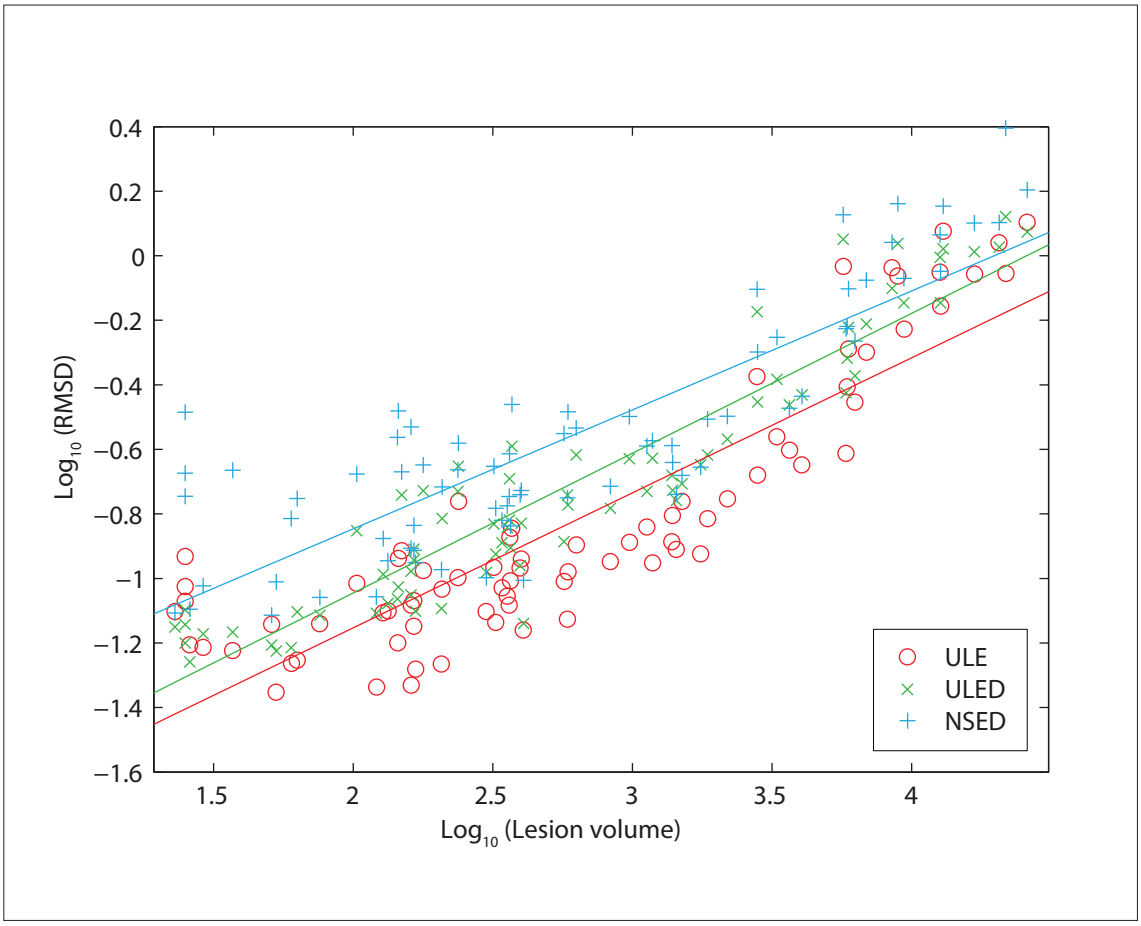
ULP vs. ULC	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	0.10	0.10	3.48	6.42E-02
Lesion volume	1	3.49	3.49	117.48	0.00E+00
Method x Lesion volume	1	1.17	1.17	39.52	0.00E+00
Error	150	4.46	0.03		

Figure 3.20 - ANCOVA comparing ULP and ULC.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation (red) against the unified segmentation-normalisation with light regularisation and cost function masking technique (blue). The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). The effect of changing the method fails to reach significance (p=0.0642).

The comparison between ULP (unified segmentation-normalisation with light regularisation) and ULC (unified segmentation-normalisation with light regularisation and cost function masking) does not show a significant difference between the methods (p=0.0642) across the lesion volume range (figure 3.20). The intersection between the 2 lines occurs at small to medium sized lesions (~315voxels), with a reduction in RMSD associated with cost function masking as lesion volume increases.

The 3 techniques using the enantiomorphic modification with the lowest RMSD were ULE, ULED and NSED (New segment with DARTEL and enantiomorphic normalisation). Figure 3.21 displays an ANCOVA analysis of the three techniques. The three methods are significantly different from each other while there is no interaction between method and lesion volume (p=0.137) with ULE having the lowest RMSD and NSED with the highest.



ULE vs. ULED vs. NSED	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	2	2.77	1.39	45.67	0.00E+00
Lesion volume	1	26.54	26.54	875.27	0.00E+00
Method x Lesion volume	2	0.12	0.06	2.01	1.37E-01
Error	227	6.95	0.03		

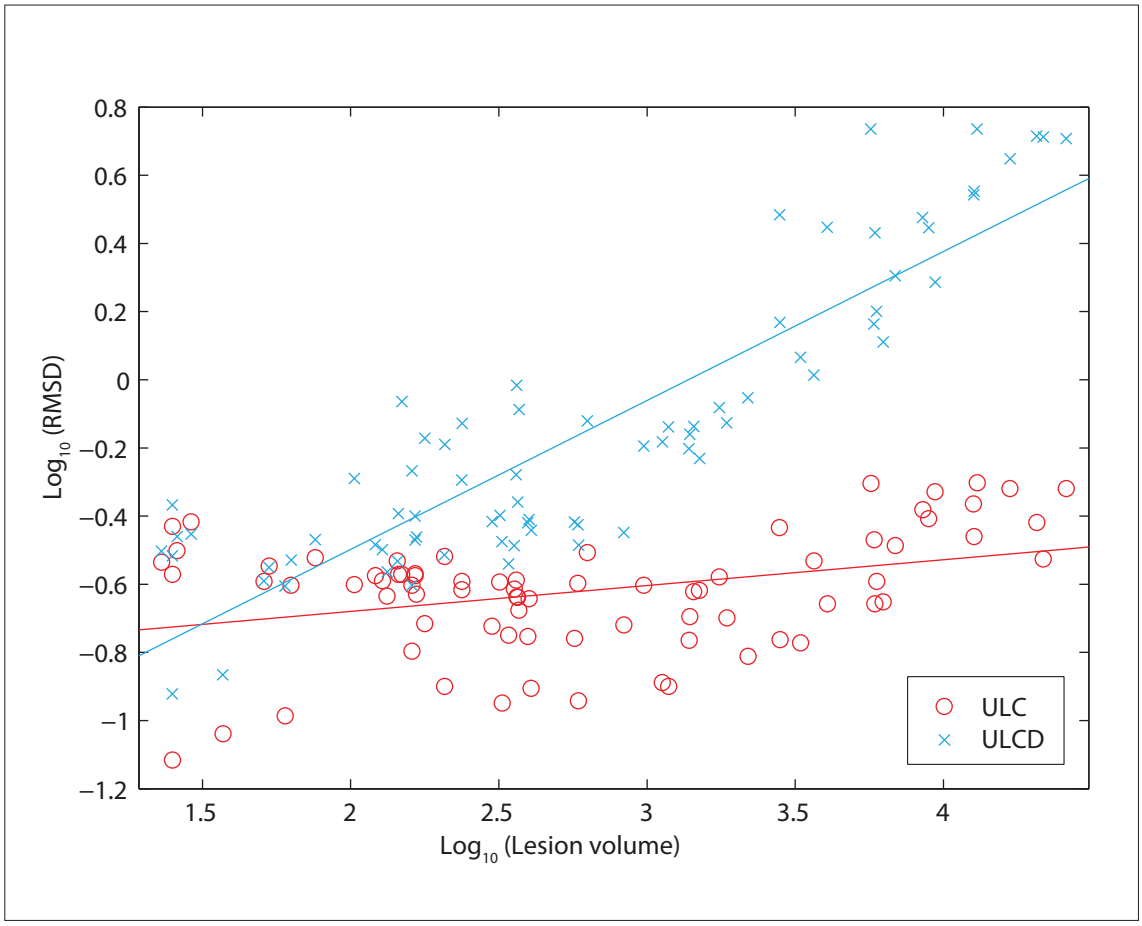
Figure 3.21 - ANCOVA comparing ULE, ULED and NSED.

ANCOVA analysis comparing 3 methods that incorporate the enantiomorphic normalisation modification. The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is an effect of method and lesion volume on the RMSD, while there is no significant interaction between method and lesion volume for these three methods.

Red: Unified segmentation-normalisation with light regularisation and enantiomorphic normalisation (ULE)

Green: Unified segmentation-normalisation with light regularisation, DARTEL and enantiomorphic normalisation (ULED)

Blue: New segment with DARTEL and enantiomorphic normalisation (NSED)



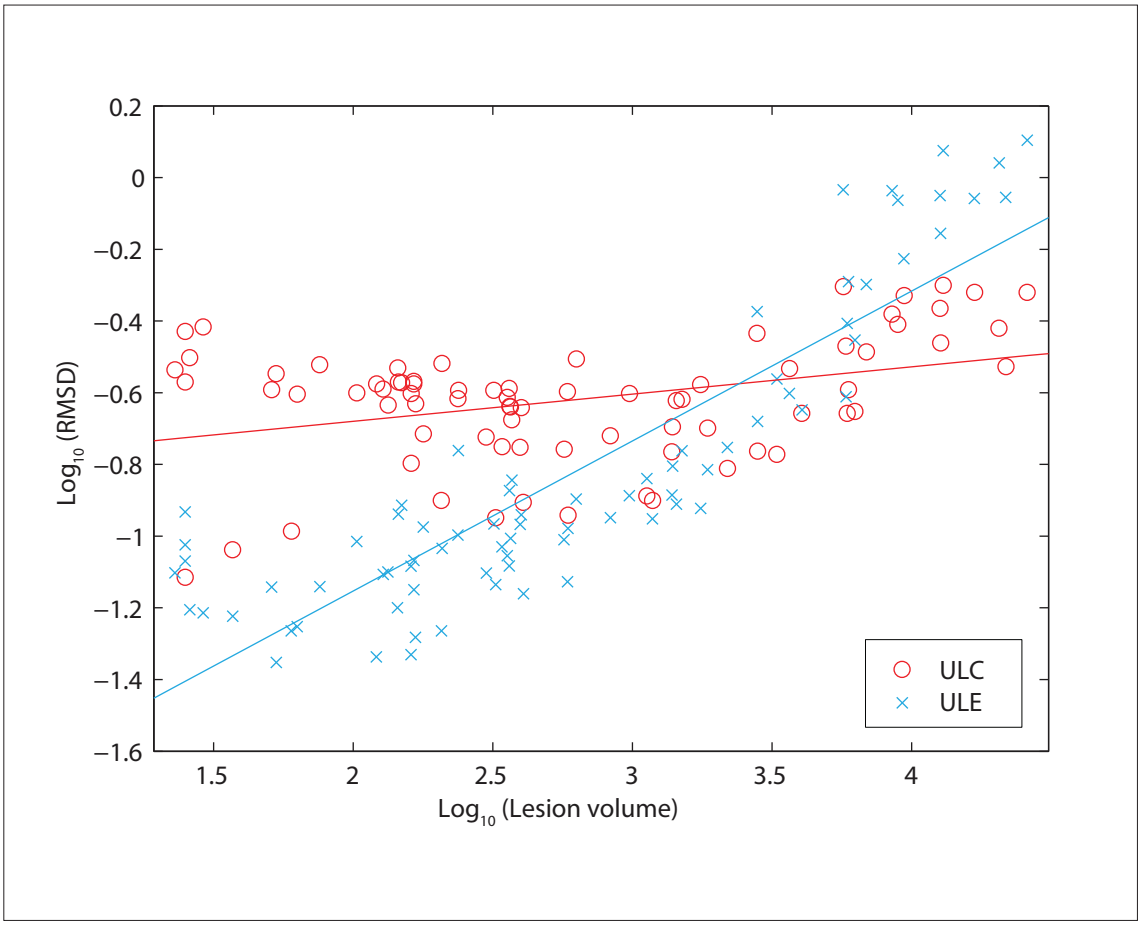
ULC vs. ULCD	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	8.33	8.33	264.18	6.72E-35
Lesion volume	1	7.03	7.03	223.00	1.79E-31
Method x Lesion volume	1	3.49	3.49	110.55	1.02E-19
Error	150	4.73	0.03		

Figure 3.22 - ANCOVA comparing ULC and ULCD.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation and cost function masking (red) against the unified segmentation-normalisation with light regularisation, DARTEL and cost function masking technique (blue). The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is an interaction between lesion volume and method, with ULCD showing a greater sensitivity to lesion volume than ULC.

Figure 3.22 compares ULC with ULCD. The interaction between method and lesion volume reaches significance (p=1.02e^{-19}) with ULCD performing worse with respects to RMSD as lesion volume increases. ULC in contrast is less affected by the change in lesion volume.

In all methods that use cost function masking, ULC has the lowest RMSD for the range of lesion volumes assessed, while for all methods that use the enantiomorphic modification, ULE has the lowest RMSD. An ANCOVA analysis comparing these two techniques, are displayed in figure 3.23. Both methods demonstrate an effect of lesion volume on RMSD (p value 1.419e^{-30}), however there is also an interaction between lesion volume and the method used (gradient$_{ULC}$=0.0759; gradient$_{ULE}$=0.4183; p value=1.2075e^{-18}). The two lines intersect at $10^{3.38}$, equivalent to 2400 voxels. Above this value ULC has a lower RMSD, while below this volume, the ULE method performs better.



ULC vs. ULE	d.f.	Sum Sq.	Mean Sq.	F	prob > F
Method	1	1.61	1.61	52.59	2.04E-11
Lesion volume	1	6.53	6.53	212.88	1.42E-30
Method x Lesion volume	1	3.13	3.13	102.19	1.21E-18
Error	150	4.60	0.03		

Figure 3.23 - ANCOVA comparing ULC and ULE.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation and cost function masking (red) against the unified segmentation-normalisation with light regularisation and enantiomorphic normalisation technique (blue). The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). There is a significant effect of method, lesion volume and interaction between the two. The 2 lines intersect at $10^{3.38}$ which is equivalent to 2400 voxels (19200mm^3).

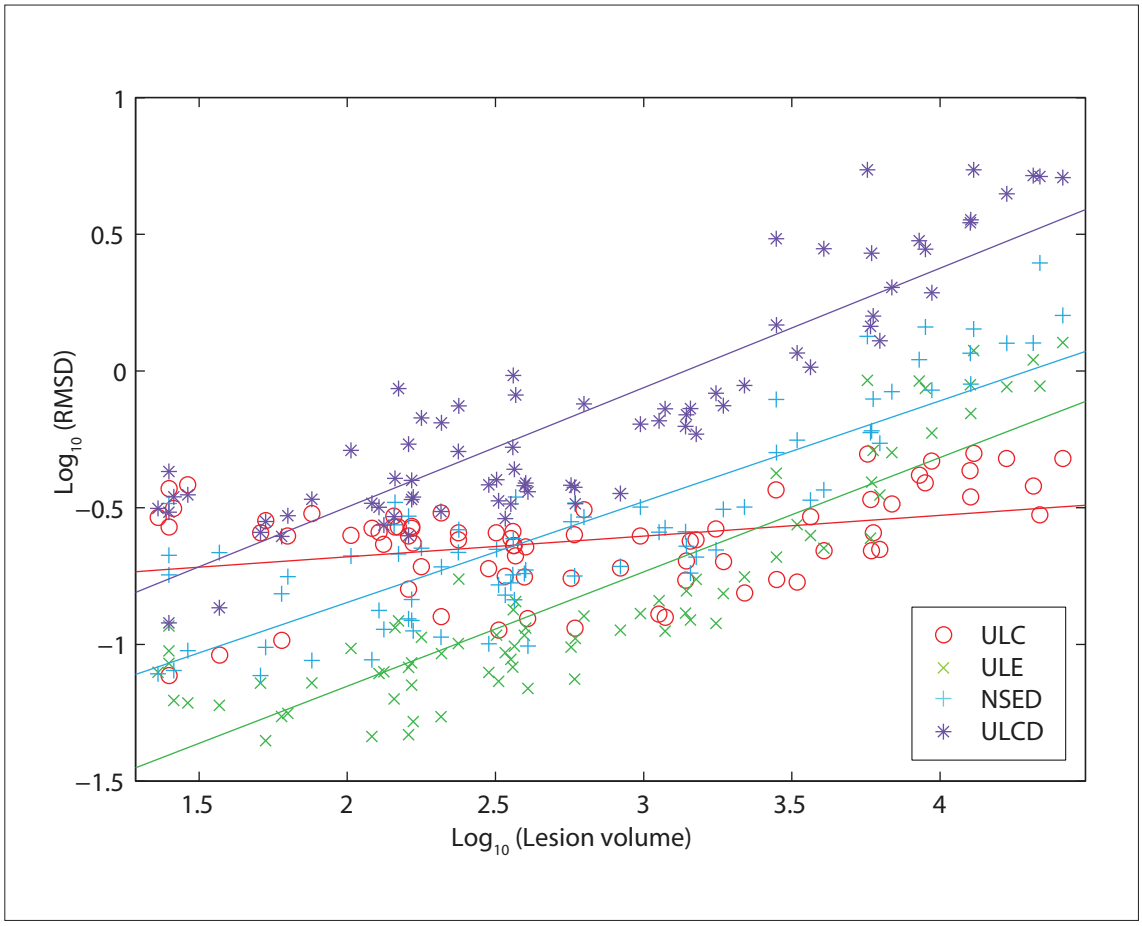


Figure 3.24 - ANCOVA comparing ULC, ULE, NSED and ULCD.

ANCOVA analysis comparing the unified segmentation-normalisation with light regularisation and cost function masking (ULC); unified segmentation-normalisation with light regularisation and enantiomorphic normalisation; New segment with DARTEL and enantiomorphic normalisation; and unified segmentation-normalisation with light regularisation, DARTEL and cost function masking. The response feature is the base 10 logarithm of the RMSD plotted against the covariate, base 10 logarithm of the lesion volume (voxels). ULE, NSED and ULCD are almost parallel with each illustrating a similar response to changes in lesion volume, whilst ULC is far less affected by the size of lesion volume displaying a more horizontal line across the lesion volume range.

Figure 3.24 displays 4 different techniques, ULC, ULE, NSED and ULCD. The effect of lesion volume on ULC is relatively small compared with the other techniques that utilise the enantiomorphic normalisation modification. The 3 enantiomorphic techniques with the lowest RMSD perform better than ULC for small lesions.

The lesion mask in native T1 space can be transformed into standard MNI space by using either the normalisation parameters from the unlesioned recipient brain or from the complementary chimeric brain. The volume ratio of unlesioned : chimeric provides an alternative measure of how much the normalisation technique is perturbed by the presence of a lesion. Figures 3.25, 3.26, and 3.27 shows that there is little variation in the ratio across the range of lesion volumes.

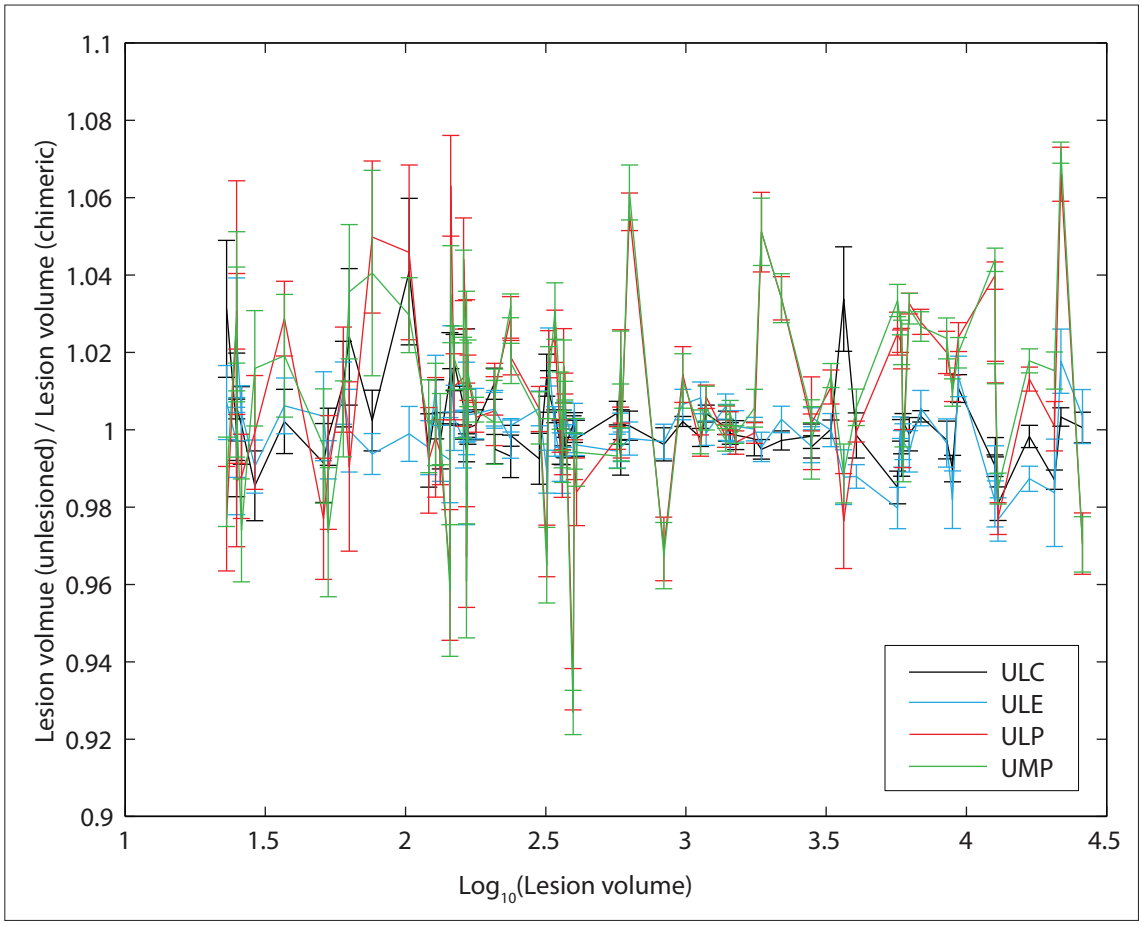


Figure 3.25 - Unified segmentation-normalisation routines:
mean volume change ratio vs Log_{10}(lesion volume).

Mean volume change ratio for different spatial normalisation techniques using the base unified segmentation-normalisation routine for each set of 11 T1 unlesioned recipient brains with the same transplanted lesion from the b1000 image. Error bars represent one standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Black: Light regularisation with cost function masking (ULC)

Blue: Light regularisation with enantiomorphic normalisation (ULE)

Red: Light regularisation (ULP)

Green:Medium regularisation (UMP)

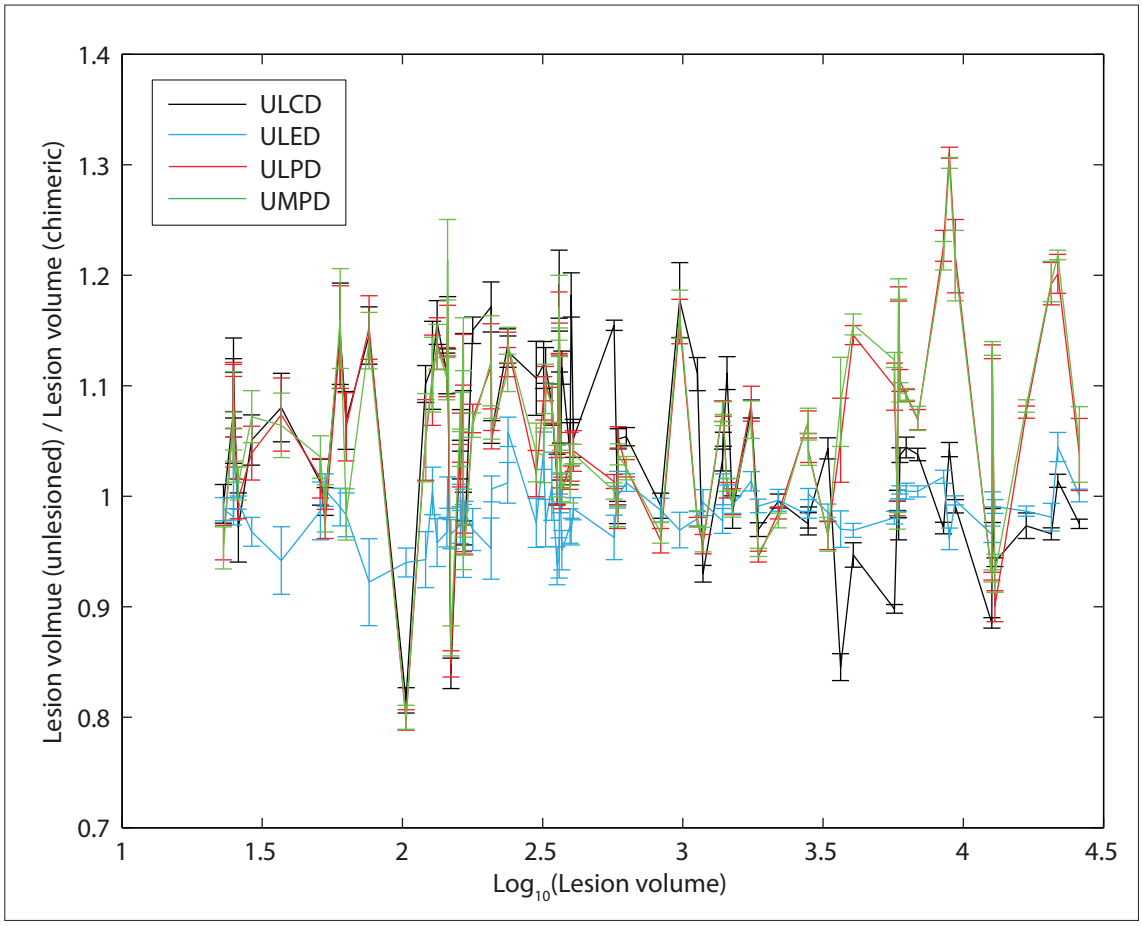


Figure 3.26 - Unified segmentation-normalisation and DARTEL routines: mean volume change ratio vs Log_{10}(lesion volume).

Mean volume change ratio for different spatial normalisation techniques using the base unified segmentation-normalisation and DARTEL routine for each set of 11 T1 unlesioned recipient brains with the same transplanted lesion from the b1000 image. Error bars represent one standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Black: Light regularisation with cost function masking (ULCD)

Blue: Light regularisation with enantiomorphic normalisation (ULED)

Red: Light regularisation (ULPD)

Green:Medium regularisation (UMPD)

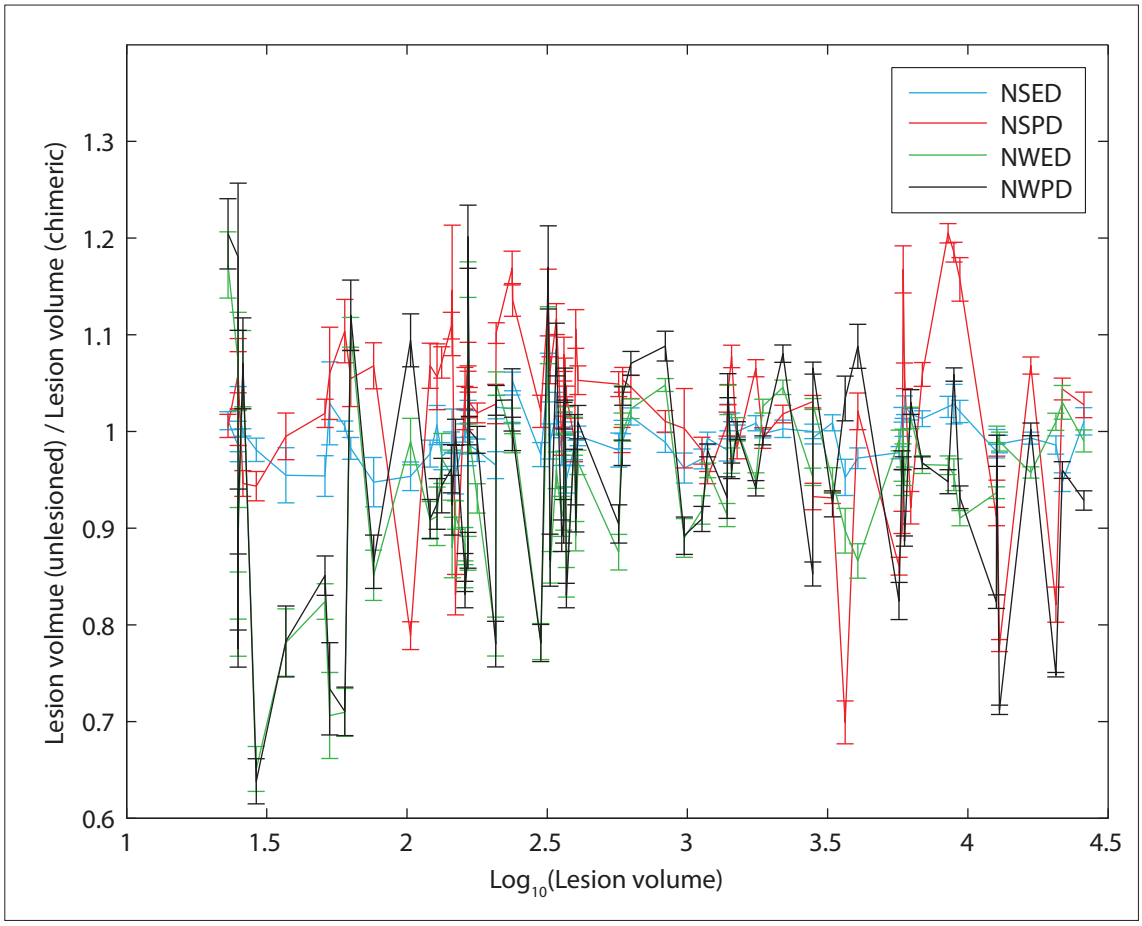


Figure 3.27 - New segment and DARTEL routines:
mean volume change ratio vs Log_{10}(lesion volume).

Mean volume change ratio for different spatial normalisation techniques using the base New segment and DARTEL routine for each set of 11 T1 unlesioned recipient brains with the same transplanted lesion from the b1000 image. Error bars represent one standard error. The lesion mask is transformed according to the spatial normalisation parameters derived from the healthy unlesioned brain and chimeric image. The volume ratio between the chimeric transformation : healthy transformation is plotted, with values greater than 1 indicating a reduction in lesion volume while values less than 1 indicate an expansion.

Blue: Enantiomorphic normalisation (NSED)

Red: No further modifications (NSPD)

Green: Extra WM/CSF prior and enantiomorphic normalisation (NWED)

Black: Extra white matter/CSF (WM/CSF) prior (NWPD)

The ANCOVA lines of ULC and ULE intersect at approximately $10^{3.38}$ (2400 voxels). Figure 3.28 plots the *p* values of the 2 sample *t*-tests comparing the 2 sets of 11 chimeric brains with the same transplanted lesion. The *y*-axis is the base 10 logarithm of the *p* value ($\log_{10}$(*p* value)) whilst the $\log_{10}$(lesion volume) is along the *x* axis. The red crosses represent the 2 sample *t*-test (right-tail) with the null hypothesis (H_0) where the RMSD of ULC is not larger than the RMSD of ULE, while the blue crosses compare ULC with ULED (2 tail) with the null hypothesis (H_0) where the RMSD of ULC is equal to the RMSD of ULED.

ULC is significantly larger than ULE with respects to RMSD, up to $10^{3.5}$ (~3000 voxels) which is consistent with the point of intersection of the 2 ANCOVA lines (*p*=0.01). Similarly there is a significant difference between ULC and ULED for all lesion volumes except for those between 10^2 – $10^{2.5}$, which also corresponds to the intersection of the ANCOVA lines for ULC and ULED (*p*=0.01).

Figure 3.29 shows the *p* values of a 2 sample *t*-test comparing ULC and ULE with respects to the volume change ratio against $\log_{10}$(lesion volume). Although the *p* value fluctuates across the lesion volume range, no significant difference between the 2 techniques is reached (*p*=0.01).

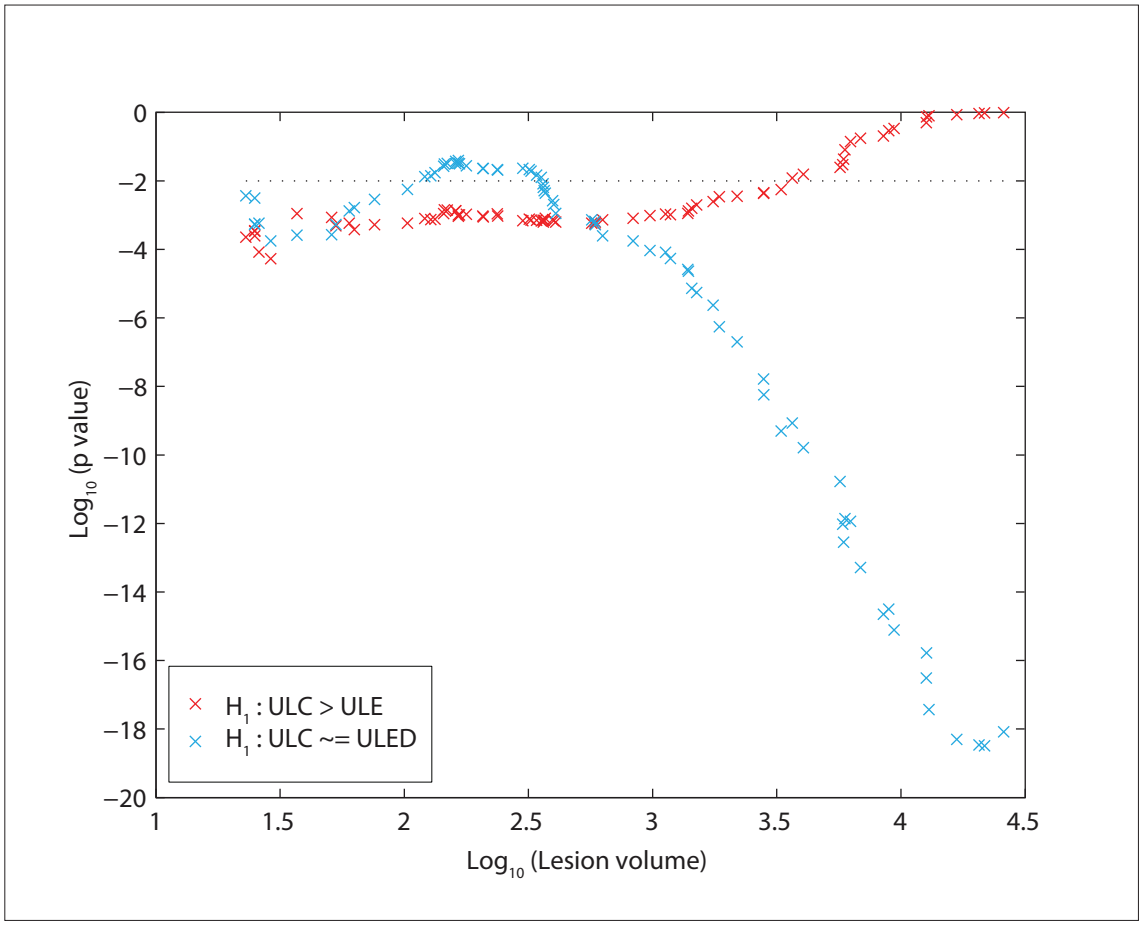


Figure 3.28 - 2 sample *t*-test assessing ULC vs ULE and ULC vs ULED.

Plot showing the base 10 logarithm of the *p*-value for 2 sample *t*-tests against the base 10 logarithm of lesion volume. The dotted line represents a *p* value = 0.01.

ULC vs ULE (red). For each set of chimeric brains with the same transplanted lesion a one tail (right) 2 sample *t*-test was performed comparing the RMSD of ULC and ULE. ULC was significantly larger, with respects to RMSD than ULE for lesions smaller than $10^{3.5}$ voxels.

ULC vs ULED (blue). For each set of chimeric brains with the same transplanted lesion a two tail 2 sample *t*-test was performed comparing the RMSD of ULC and ULED. The two techniques were not significantly different with respects RMSD for lesions between 10^{2} and $10^{2.5}$ voxels.

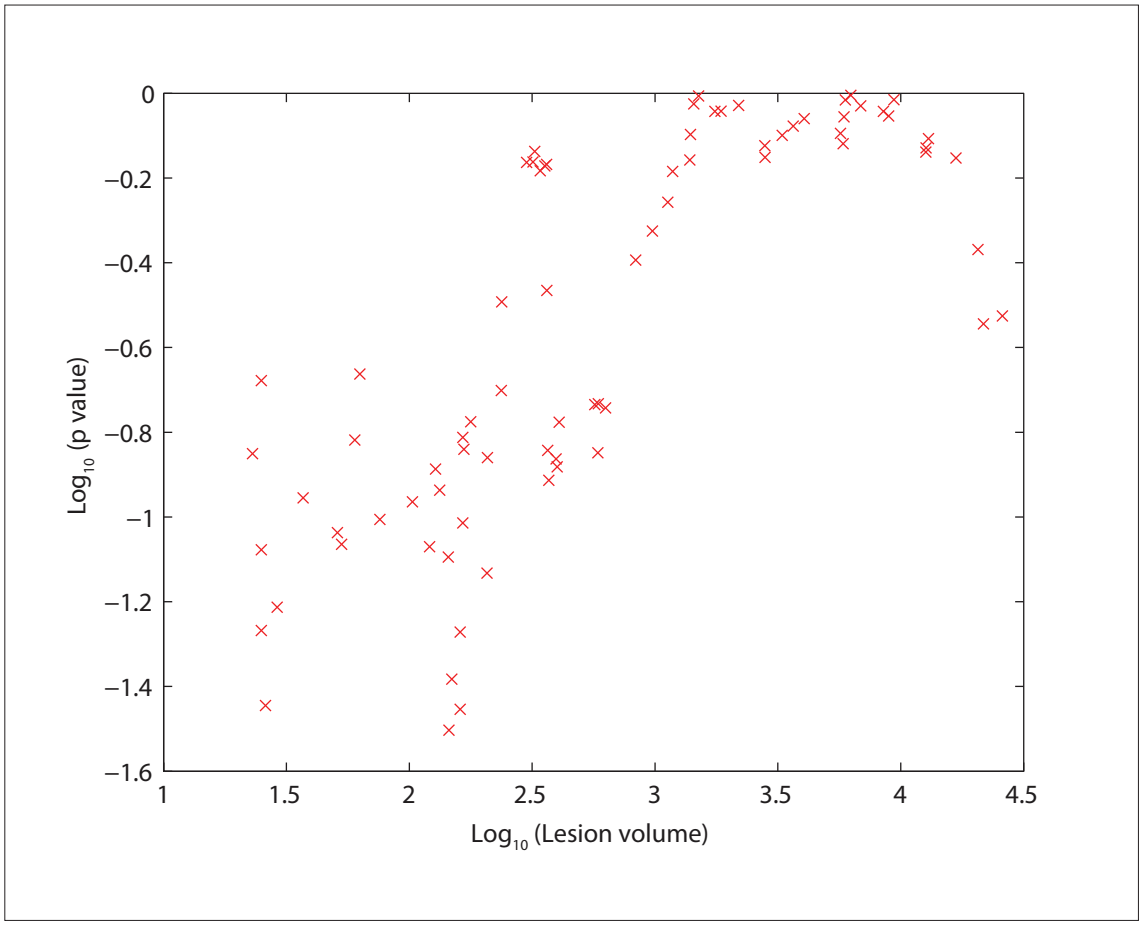


Figure 3.29 - 2 sample *t*-test assessing ULC vs ULE.

Plot showing the base 10 logarithm of the *p*-value for a 2 tail, 2 sample *t*-tests against the base 10 logarithm of lesion volume. The two techniques assessed were ULC and ULE with respects to the volume change ratio. No significant difference was found across the entire lesion volume range.

3.4 Discussion

In the preceding series of simulations, a total of 12 different spatial normalisation techniques were compared. The performance of each technique was assessed on 847 chimeric brain volumes, derived from 11 healthy T1 brain volumes and 77 lesions, manually extracted from b1000 diffusion weighted volumes, resulting in a total of 10164 normalised brains.

Each normalisation method is assessed by comparing the deformation fields generated when transforming a brain volume into normalised space in the presence and absence of a lesion (Brett et al., 2001; Nachev et al., 2008; Ripollés et al., 2012). Therefore it is not an assessment of how well a method spatially normalises a brain, but to what extent is the method perturbed by the presence of a lesion. Conceptually, if the normalisation result is greatly affected by a lesion, the transformation derived will probably be worse than in the absence of the lesion since the latter case would be considered the optimal situation.

In this study a set of artificially lesioned brains (chimeric brains) was created so that the state of the brain in the lesioned and unlesioned state was known. The analysis grouping the chimeric brains according to their common background (recipient) brain demonstrated very little variation in both the volume change ratio of the lesion and the root mean squared difference (RMSD).

Previous work by Nachev et al (Nachev et al., 2008) demonstrated a relationship between the RMSD and the size of the lesion. This relationship could be transformed into a linear relationship by assessing the $\log_{10}$(RMSD) with the $\log_{10}$(lesion volume).

In the presence of DARTEL (Ashburner, 2007), the addition of the enantiomorphic modification (Nachev et al., 2008) to either the unified

segmentation-normalisation or New segment base methods (Ashburner and Friston, 2005) resulted in a downward shift in the lines. Both the ULED and NSED (with enantiomorphic correction) performed better across the entire lesion volume range compared with ULPD and NSPD (without enantiomorphic correction) respectively. Moreover there was no interaction found in either of the two parings (figures 3.16 and 3.17). DARTEL uses flow field across the entire brain, and imposes certain restrictions to ensure reasonable contours of these paths. This result suggests that the transplanted contralesional signal intensities (enantiomorphic information) has greater utility than the lesion tissue with DARTEL. Despite ultimately being abnormal for its location, it is more representative of what might have been present prior to the injury.

If the availability of information about the entire brain volume influences the performance of normalisation techniques using DARTEL, the absence of information should result in a poorer outcome, with a larger RMSD. This situation is examined in figure 3.18 where the unified segmentation-normalisation with light regularisation and DARTEL (ULPD) is compared with the technique that excludes the information in the lesion voxels, the unified segmentation-normalisation with light regularisation, DARTEL and cost function masking (ULCD). As predicted the exclusion of the lesioned voxels results in an increase in RMSD across the entire lesion volume range.

Previous work assessing the benefit of the enantiomorphic modification demonstrated a decreasing advantage with increasing lesion volume (Nachev et al., 2008). This pattern is supported in figure 3.19 where ULP is compared with ULE. The two lines begin to converge with large lesions (~5500 voxels, 44,000mm^3). Theoretically as the lesion size decreases the difference in RMSD between the 2 techniques should asymptotically approach zero. Figure 3.19 does suggest that this feature is about to emerge at the lower end of the lesion volume range.

In this dataset no significant difference was found between the unified segmentation-normalisation with light regularisation and cost function masking (ULC) and ULP. This outcome is echoed in the work by Crinion et al who proposed the unified segmentation-normalisation technique performed similarly to the technique with cost function masking (Crinion et al., 2007). Their work was in response to that of Brett et al who advocated the use of cost function masking, who proposed excluding the abnormal voxels from interfering with the normalisation process. This was later supported from work by Andersen et al. who suggested even coarse lesion masking can improve the normalisation result (Andersen et al., 2010; Brett et al., 2001). Although no significant difference was found between ULP and ULC, figure 3.20 would suggest that cost function masking performs better as the lesion volume increases and supports the use of cost function masking for medium to large lesions.

Unusually the unified segmentation-normalisation with light regularisation and enantiomorphic normalisation (ULE) has a lower RMSD across the entire lesion volume range than either ULED or NSED which also use DARTEL. The lines in figure 3.21 do not show a significant interaction between lesion volume and method. This difference may be secondary to DARTEL being more sensitive to the presence of abnormal signal. As illustrated above, as the quality of the information in the lesioned area improves the RMSD decreases. In essence the transplanted enantiomorphic information is still abnormal for its location and the trade-off between directly managing this abnormality is not overcome by the improvement offered by the enantiomorphic correction. One explanation is that DARTEL makes use of the deformations present in the surround extra-lesional tissue caused by oedema to compensate for the lesioned voxels. Unfortunately this extra-lesional detail is not available with the artificially lesioned (chimeric) brains.

The 2 best performing techniques are the unified segmentation-normalisation with light regularisation and cost function masking (ULC) and unified segmentation-normalisation with light regularisation and enantiomorphic normalisation (ULE). For small to medium lesions ULE has a lower RMSD than ULC with the 2 lines intersecting at to 2400 voxels (19,200mm^3). It is not entirely clear whether the two techniques are significantly different for large lesions. This result supports the finding by Nachev et al, who identified a limit to the improvement available from enantiomorphic normalisation asymptotic to cost function masking (Nachev et al., 2008).

The 3 best performing techniques with respect to RMSD are ULC, ULE and ULED. Figure 3.28 examines each lesion set in turn and performs a 2 sample test comparing: ULC vs. ULE and ULC vs. ULED. The mean RMSD for ULE is significantly different to ULC up to 3000 voxels (p=0.01). This is consistent with the ANCOVA analysis assessing ULC with ULE which demonstrated an intersection between the 2 lines around the same lesion volume. When comparing ULED with ULC, a similar reflection of the ANCOVA analysis is found, with a significant difference (p=0.01) found up to 100 voxels and then from 300 voxels onwards (figure 3.24).

A couple of groups have stated that the reduction in RMSD needs to be considered in combination with the percentage volume change of the lesion after normalisation, since many algorithms will inevitably attempt to shrink the lesions to minimise the signal difference between the lesioned brain and reference brain. Consequently past studies have shown a decrease in lesion volume (compared with CFM) especially with methods that utilise DARTEL. The spatial normalisation parameters derived from the unlesioned T1 brain were applied to the lesion mask and the number of voxels occupied by the transformed lesion was used as the numerator, whilst the number of voxels occupied by the lesion mask transformed using the normalisation parameters

derived from the chimeric brain were used as the denominator. The deviation of this ratio away from one can then be assessed between techniques. Similar to Ripollés et al and Andersen et al, DARTEL and unified segmentation-normalisation with light regularisation (ULP) resulted in a reduction in final lesion volume after transformation. Interestingly, cost function masking also resulted in shrinkage in this study, though the amount was less. However the opposite was noted for all the methods that used an enantiomorphic modification, with lesions generally undergoing expansion after transformation.

The effect of lesion volume on this ratio can be examined. Figure 3.29 shows that there is no significant difference in the absolute deviation from 1 of this ratio between ULC and ULE across the entire lesion volume range (p=0.01). It should be noted that at very small lesion volumes significance is almost reached. The act of enantiomorphic correction is still replacing abnormal signal with signal that is ultimately abnormal for its location. Therefore at very small lesion volumes, the transplanted signal may in fact mimic a lesion more than normal tissue and masking out the area would be of greater benefit.

For both RMSD and volume change ratio after normalisation, there was little variation across the different backgrounds.

3.5 Conclusion

The preceding study has shown that the enantiomorphic modification to the unified segmentation-normalisation with light regularisation method (ULE) to perform the best out of all the techniques assessed. Importantly its performance was better with respect to the root mean squared difference compared with the unified segmentation-normalisation with light regularisation and cost function masking (ULC) at small to medium sized lesions. This advantage was seen in lesion sizes up to 3000 voxels (24000mm^3) after which the two methods had comparable results. No significant difference was found between these two techniques with large lesion volumes. In addition no significant difference between ULC and ULE was found with respect to the amount lesion volume change after normalisation. Consequently in the presence of a small to medium sized unilateral lesion, it is recommended to perform spatial normalisation with the enantiomorphic modification.

4 Lesion function mapping using mass univariate techniques

4.1 Introduction

The beginnings of functional localisation within the brain can be traced back to work by Paul Broca and Carl Wernicke in the mid 19th Century. Through careful clinical observations of patients and subsequent post mortem examination of the patients' brains, Broca correlated the ventro-posterior region of the frontal lobe to speech production, whilst Wernicke related the left posterior, superior temporal gyrus to language comprehension (Broca, 1861; Wernicke, 1874). With these discoveries, they revealed the human brain to possess a functionally specialised architecture. Their conclusions were based on the visual assessment on a number of brains, noting the size and location of injury and documenting regions of commonality. Their approach was, however, hindered by the small available sample size of suitable patients and the moral requirement to await their timely passing. The arrival of computed tomography (CT) and Magnetic Resonance Imaging (MRI) opened a new dimension to lesion function mapping (Damasio and Damasio, 1989; Rorden and Karnath, 2004). With these tools, brain injury could be visualized, quantified and monitored *in vivo*. This has not only increased the number of patients available to study, but also increased the spatial resolution of analysis, improving temporal resolution and providing a means to collect suitable control subjects. As a result over the last century and a half such studies have been critical to identifying the distinctive neural substrates of language (Bates et al., 2003; Dronkers et al., 2004), memory (Scoville and Milner, 1957), emotion (Adolphs et al., 1995; Calder et al., 2000), attention (Egly et al., 1994; Karnath et al., 2004; Mort et al., 2003) and intelligence (Gläscher et al., 2009).

Functional Magnetic Resonance Imaging (fMRI) has become a popular tool for investigating the functional architecture of the brain. The technique exploits the different magnetic properties of oxygenated and deoxygenated blood to generate a blood oxygen level dependent (BOLD) signal. The crux of fMRI studies is the association between increased neuronal activity and oxygen requirements. In this technique's favour is the spatial resolution of 1-5mm (Menon and Kim, 1999), and a temporal resolution of seconds. Importantly fMRI permits the observation of the brain but cannot interfere with brain function, therefore necessitating the manipulation of the experimental design to generate contrasts in the BOLD signal. As a consequence, fMRI provides a powerful approach to hypothesising putative critical regions but has greater difficulty in testing them (Aue et al., 2009). The necessity of a brain region for a putative function – arguably the strongest test – can only be established by showing a deficit when the function of the region is disrupted.

Inactivating brain areas experimentally cannot easily be done in the human. The technique of transcranial magnetic stimulation (TMS) provides a non-invasive approach to temporarily disrupting a region of the brain. Although temporally its resolution is high, in the order of milliseconds, its spatial resolution is very limited, particularly centripetally, as the effects of stimulation are restricted to superficial cortical regions (Epstein et al., 1990; Rudiak and Marg, 1994; Walsh and Cowey, 2000; Zangen et al., 2005).

The only comprehensive means of establishing necessity is therefore the study of patients with naturally occurring focal brain lesions (Rorden and Karnath, 2004). The majority of these studies have involved a cohort of patients with statistics performed on the group rather than the individual. As a consequence the brain volumes must all be brought into spatial register, by transforming each image so as to align homologous regions between images, enabling point-by-point anatomical comparisons to be made across the cohort.

When the brain is damaged by a focal pathological process the pattern of damage generally bears no relation to the underlying functional architecture. Critically, the scale of damage in clinical cases (typically 10^{-2}m) is substantially greater than the scale of functional organisation suggested by the spatial heterogeneity of individual neuronal responses ($<10^{-3}$m). To make population-level inferences about the functional role of a given part of the brain from lesion data, previous studies have therefore relied on comparing sets of patients with large, inevitably overlapping, lesions to identify a critical locus much smaller than each individual lesion (e.g. (Karnath et al., 2004)). The comparison is made by applying a statistical test point-by-point for each part of the brain, discretized at some convenient spatial resolution, where each point is treated independently (Bates et al., 2003; Karnath et al., 2004). This mass-univariate approach assumes that the other, apparently non-critical areas damaged in each patient do not distort the localisation of the one, critical area that is common to them all: in short, that any spatial correlations in the pattern of damage within each patient are well-behaved. However, it is these hidden systematic biases in the natural patterns of damage that may invalidate our anatomical inferences. In relation to overwhelmingly the commonest lesion type used in such studies – ischaemic vascular – the assumption is fundamentally unsafe because the architecture of the vascular tree is highly stereotyped across individuals.

To illustrate the potential consequences of this mass univariate approach in the presence of hidden structure in the data, consider the two-dimensional synthetic example in figure 1.6, where damage to any part of area *A* alone may disrupt a putative function of interest but *B* plays no role in this function of interest. If the lesions used to map the functional dependence on *A* follow a stereotyped pattern where damage to any part of *A* is systematically associated with collateral damage to the non-critical area *B*, both areas may appear to be significantly associated even if *B* is irrelevant to the function of

interest. Crucially, if the pattern of the lesions within each patient is such (for reasons to do with factors unconnected to function) that the spatial variability of damage to *B* is less than to *A*, *B* will not only be erroneously determined to be critical but will have a higher significance value for such an association than *A*. The apparent locus of a lesion-function deficit will therefore be displaced from *A* (the true locus) to *B*. A hidden structure in the pattern of damage – hidden because it is apparent only when examining the pattern as a whole, in a multivariate way – distorts the spatial inference.

Mass-univariate techniques that perform a statistical test at each point location ignore the spatial dependencies of the data. As illustrated in figure 1.6, merely increasing the dataset available to the technique cannot overcome this problem; it can only amplify the biases. In other words, any error resulting from such systematic biases can neither be corrected through redundant effort with the same technique, nor through another technique, for none has comparable inferential power.

The question of whether or not such biases exist is therefore of crucial importance to systems level neuroscience. Examining a very large set of vascular lesions – overwhelmingly the commonest type used in mapping studies – here we show that systematic biases exist and that they lead to substantial mislocalisation even where the relation between anatomy and function is asymptotically simple.

In the following chapter a cohort of 581 vascular injured brain volumes are used in a series of simulations utilising the mass univariate technique to try and quantify the magnitude of the mislocalisation across the brain.

4.2 Methods

The aim of this experiment is not to investigate the locus of a particular function, but to examine the potential systematic bias inherent in previous lesion function mapping methods. In both cases we need a cohort of brain volumes with their inherent injury segmented to create a binary mask upon which the inferential statistics are performed. All simulations were performed using MATLAB (The MathsWorks Inc.) on an Intel i7 920 2.66GHz processor, with 6GB RAM and 120GB solid state drive (SSD).

4.2.1 Imaging

A total of 581 brain scans were collected from the National Hospital for Neurology and Neurosurgery for this series of simulations.

All brain scans possessed a vascular lesion and were performed on a GE Genesis Signa 1.5 Tesla (1.5T) MRI scanner in a single session for each patient. The scans collected were the axially acquired T2-diffusion weighted echoplanar images (b0 and b1000 sequences) sampled at 1mm x 1mm x 6.5mm (TR=10000ms, TE=104.9ms, Columns=256, Rows=256, Slice thickness=5mm). The mean age was 62.3 years (standard deviation (SD) = 17.8 years), and the proportion of males was 0.547.

4.2.2 Image preprocessing

For each patient the corresponding b0 and b1000 images were reconstructed into a 3 dimensional volume using xmedcon (Nolf et al., 2003). The b1000 volume was rigidly co-registered to the b0 volume using the co-registration function in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and its default parameter settings (the specific parameter settings are listed in appendix B). The b0

volume was subsequently spatially normalised into MNI space (ICBM space – European brains – template) using the combined segmentation-normalisation method in SPM8 (Crinion et al., 2007), with the derived normalisation parameters applied to the co-registered b1000 sequence. Again the default parameter settings for the segmentation-normalisation method provided by SPM were used, except for the interpolation setting in the writing option being set to a 6^{th} degree spline.

A binary mask for each patient was created via the lesion segmentation using the zeta anomaly detection method described in §2 (Mah et al., 2012). The resultant masks were sliced at 2mm isovoxels and occupied a volume measuring 91 x 109 x 91 voxels. Figure 4.1 displays the volume coverage of the 581 different lesions.

Although there is evidence to show an increased preponderance to injury to the left hemisphere secondary to the anatomical arrangement of the aortic arch, it is generally believed the distribution within each hemisphere is similar (Naess et al., 2006). Consequently to increase the density of lesion coverage per voxel, all unilateral left sided lesions were reflected in the mid-sagittal plane. For bi-hemispheric lesions, the hemisphere with the larger lesion load was treated as the dominant side, with left hemisphere dominant bi-hemispheric lesions reflected in the mid-sagittal plane. Figure 4.1 displays the volume coverage of the 581 binary masks for the whole brain, while figure 4.2 shows the volume coverage of the right hemisphere after collapsing the binary masks onto a single hemisphere.

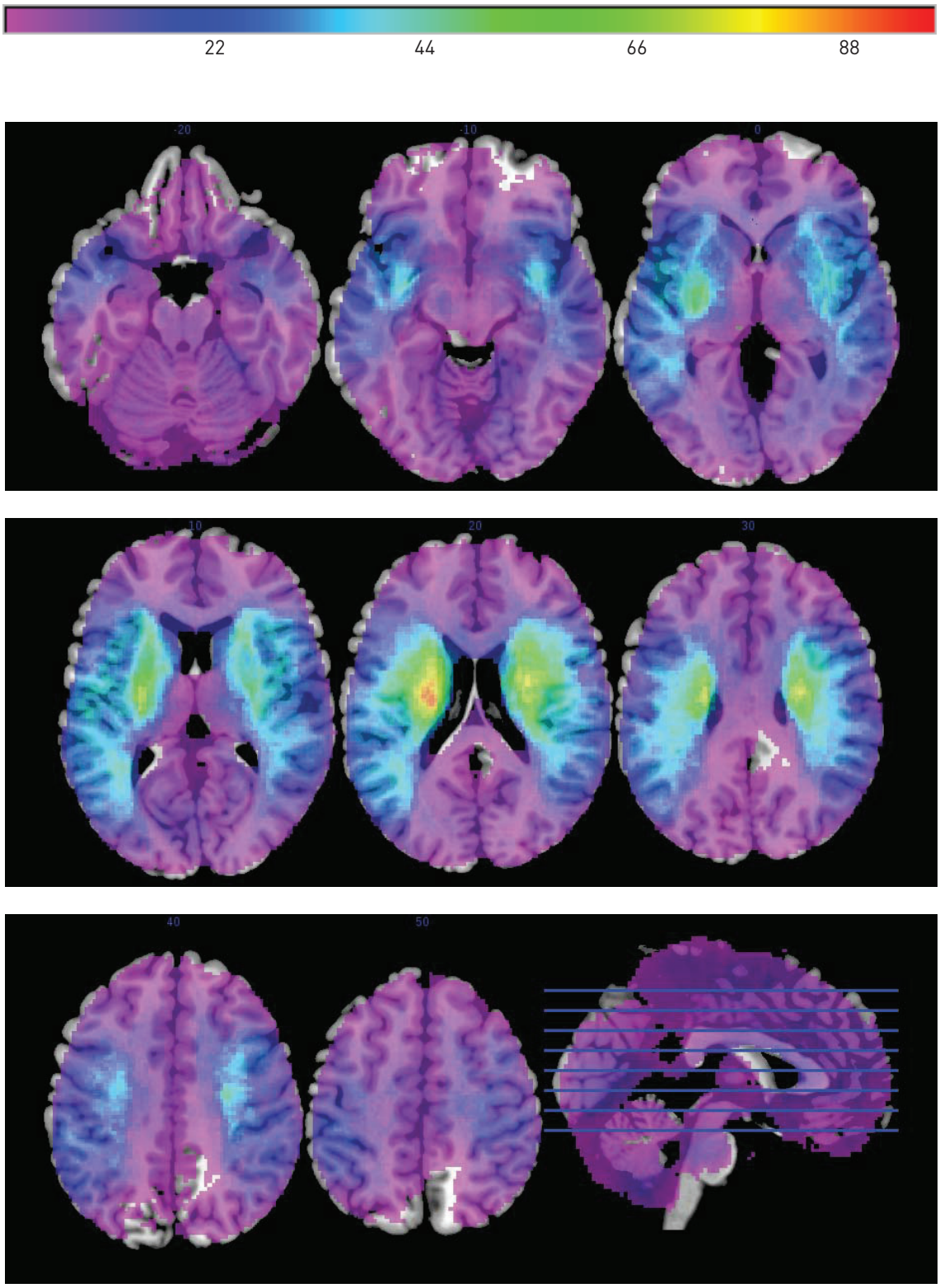


Figure 4.1 - Lesion overlay map of the 581 lesion masks.

The axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. The peak overlay value is 96. The distribution of lesion injury is symmetrical across the mid-sagittal plane.

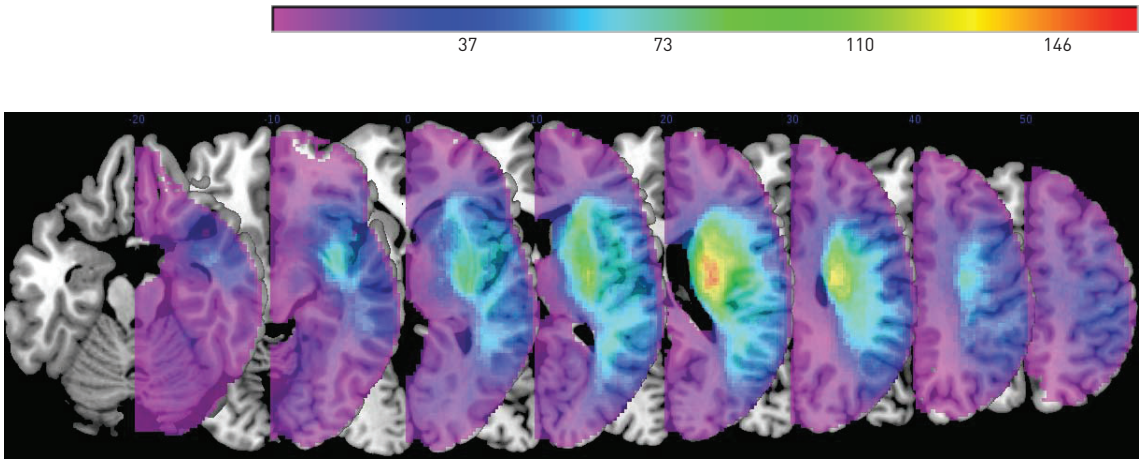


Figure 4.2 - Lesion overlay map of the 581 lesion masks collapsed on the right hemisphere.

The axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. The peak overlay value is 160.

4.2.3 Data analysis

For the subsequent experiments only the within brain voxels were used. The brain mask template from SPM8 was thresholded at less than 0.8, and greater than 0.4 to identify the within brain voxels. Since the lesions have been collapsed onto the right hemisphere, only those voxels in the right hemisphere were then identified and used to create a 581 by 105266 data matrix.

4.2.3.1 Visualisation of a high dimensional data set in two dimensions

To reveal a hidden structure that may exist within the high dimensional data set, the 581 binary masks were processed through a dimensionality reduction process with each voxel within the volume treated as a single dimension whose value was a boolean representing either healthy or lesioned tissue.

A coarse dimensionality reduction step where all voxels that were not affected by more than 3 lesions in the 581 dataset were excluded from the analysis, thereby reducing the total number of dimensions to 90469.

Similar to van der Maaten and Hinton (2008) the data was first processed with Isomap (Tenenbaum, 2000) to reduce the dimensionality of the data; improve computation speed and suppress noise. The 581 dataset with 90469 dimensions was processed with Isomap. The Dijkstra algorithm was used to compute the graph distances and the entire 581 by 581 distance matrix was utilised. The standard K method was used with the *k* nearest neighbour (*kNN*) set to 7, to redistribute the 581 masks into a 30 dimensional volume space.

The dimensionally reduced dataset was passed to tSNE to embed the dataset into a 2 dimensional space to facilitate visualisation in a 2 dimensional space. The perplexity was set to 5. Perplexity is a tuning parameter for tSNE which can be considered a smooth measure of the effective number of neighbours.

Although the performance of tSNE is fairly robust under different perplexity settings, generally larger or denser data sets require larger perplexity values (van der Maaten and Hinton, 2008). The 2 values associated with each binary mask were then used as a co-ordinate to plot a point in two dimensional Cartesian space.

4.2.3.2 Mass univariate inference

To determine how the conventional methodology of lesion mapping affects the fidelity of the result the true neural (spatial) dependence of any putative function must first be known. This was achieved by specifying a set of hypothetical models which would be regarded as the "ground truth".

The first set of hypothetical models was the dependence of a putative function of interest on a single voxel. Each voxel within the brain volume would be considered in turn (voxel-wise) as the critical locus. The second set of hypothetical models examined the dependence of a function on a region of interest, in this case a single Brodamann area.

4.2.3.2.1 Voxel-wise simulations examining the dependence of a putative function of interest on a single voxel

The simplest possible model is the single voxel dependence of a putative function of interest. Only voxels that had more than 3 binary masks affecting it were included in the analysis. For each of the 90469 models corresponding to each voxel location hit at least 4 times in the dataset, a single voxel location in the brain was taken as being critical to a hypothetical function. Each of the 581 scans within the dataset was then classified as either "affected" or "unaffected" depending on whether or not that voxel fell within the lesion present in the scan (figure 4.3). This provided a simulated "ground truth" label

for the model, splitting the set into 2 groups, just as if they were two patient groups differing in behaviour.

After all 581 brain scans have been classified as either “affected” or “unaffected” based on the single voxel dependence model, a univariate test statistic, in this case a Fisher’s exact test, is performed at each voxel location to generate a brain map of *p* values for that specific single voxel model (§ 4.2.3.2.1.1). This *p* value brain map is then thresholded and the centre of mass for this thresholded volume derived. The amount of mislocalisation is therefore the displacement from the true voxel location to the method’s derived centre of mass. The process is then repeated across the entire brain for each single voxel model.

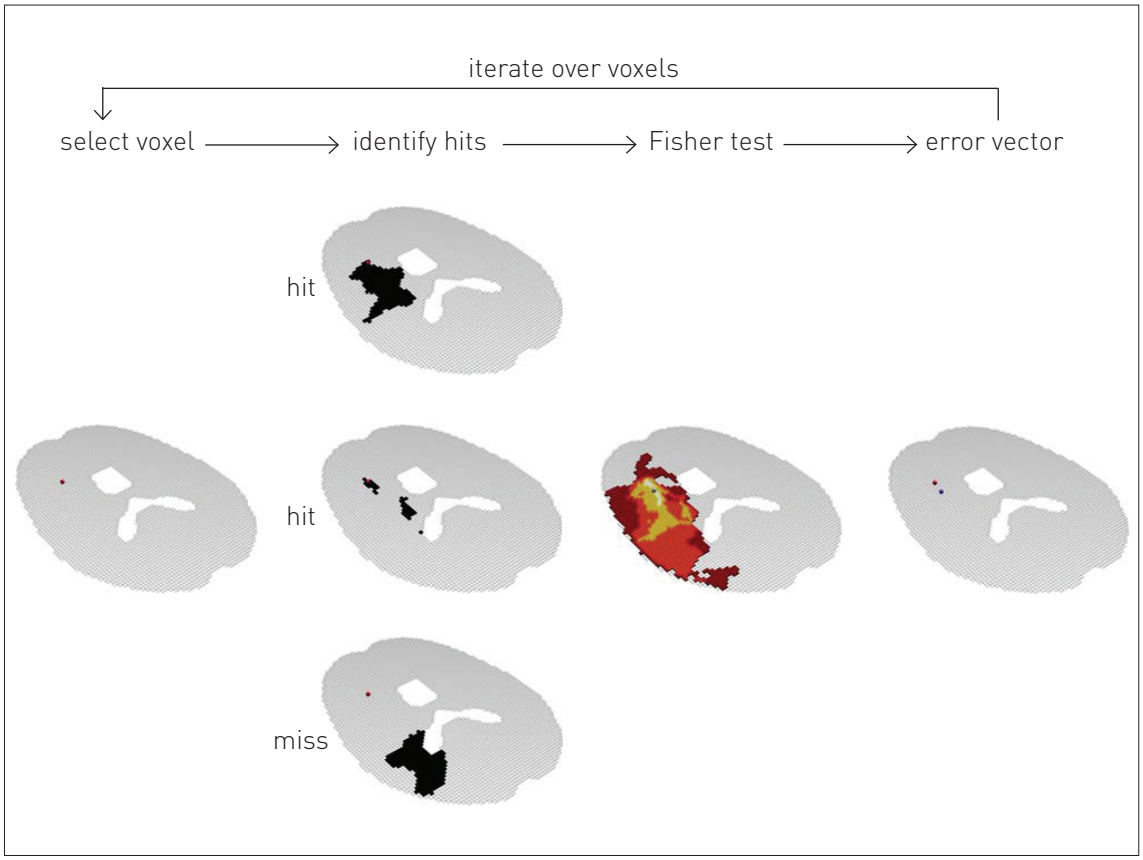


Figure 4.3 - Flow diagram illustrating the single voxel dependence of a putative function of interest simulation.

A putative function of interest is treated as dependent on a single (true) voxel location (column 1). Next every binary mask in the cohort of 581 lesioned brains is classified as either “affected” or “unaffected” depending on whether the critical voxel lies within the lesion volume of the binary mask (column 2). A statistical test (Fisher’s exact test) is performed at every voxel location to generate a brain map of *p*-values (column 3). The *p*-value brain map is then thresholded and the centre of mass for this thresholded volume is derived. The amount of mislocalisation is therefore the displacement from the true voxel location to the method's derived centre of mass (column 4). This process is repeated across the entire brain to generate a displacement vector for each voxel in the brain (voxel-wise).

4.2.3.2.1.1 Statistical analysis

At each voxel location a 2 x 2 confusion table was constructed where the 2 binary states were lesioned or healthy, and affected or unaffected, with the frequencies derived from the 581 scans. A mass-univariate analysis was performed at each voxel location, treating each voxel independently of every other, producing a voxel wise *p* value map across the brain testing the null hypothesis that the voxel is unrelated to the group label. The statistical test was Fisher's exact test (Fisher, 1970). The resultant *p* map was thresholded at $p<0.01$, Bonferroni corrected for multiple comparisons (0.01/105266). This process was repeated for every one of the 90469 voxels in the dataset that was affected in at least 4 of the set of 581 scans.

4.2.3.2.1.2 Calculation of the vector displacement

The cluster of significant voxels isolated after thresholding the resultant *p* map was reduced to a single point locus by calculating the centre of mass for the cluster. As we are working with binary masks derived from thresholding the *p* map, we can attribute a weight of 1 or 0 depending on the state of the voxel, "also lesioned" and "healthy". Utilizing the fact that we are in standardized normalised space, a common origin can be specified and the moment of each voxel from the origin can be calculated for each axis (figure 4.5). The 3 derived values then represents our new co-ordinates for the cluster's centre of mass.

$$centre\ of\ mass_{axis} = \frac{\sum m_i r_i}{\sum m_i}$$

m is the mass of the voxel

r is the distance of the voxel from the origin

The displacement from the label-defining voxel to the cluster's centre is calculated to provide a vector indicating the direction and magnitude of error introduced by the mass univariate technique.

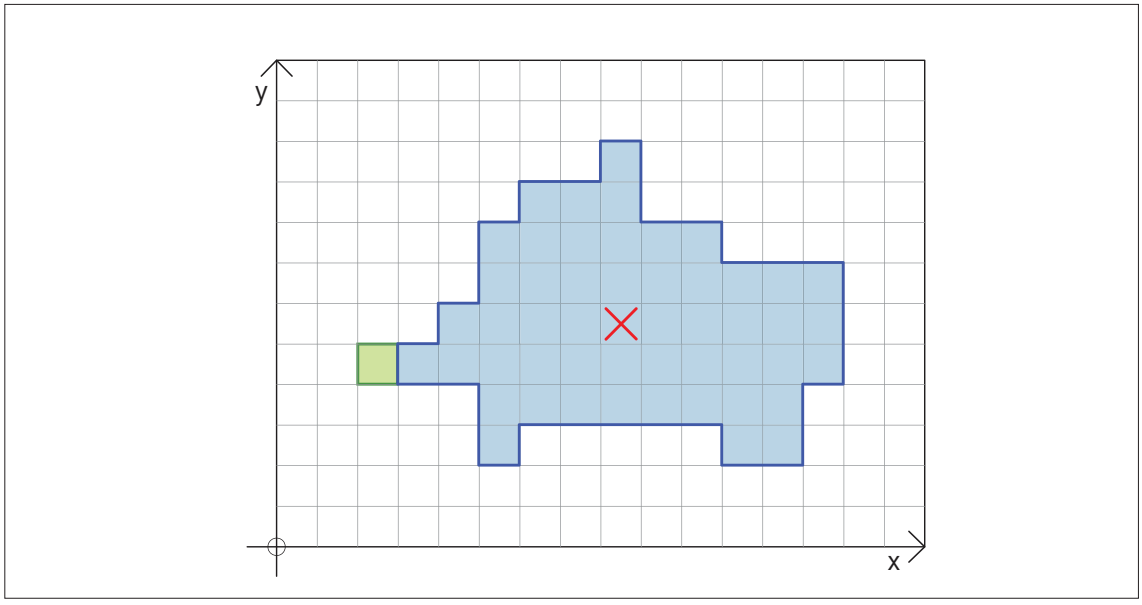


Figure 4.4 - Illustration for the calculation of the centre of mass.

To calculate the centre of mass of the injured region, each lesioned voxel is attributed a weight of unit one and the moment calculated from a common origin. In the 2 dimensional example above the moment in all axes are calculated. Therefore the moments for the green voxel would be 3 and 5 for the *x* and *y* axes respectively. The centre of mass for the coloured region is therefore (9, 6) indicated by the red cross.

4.2.3.2.1.3 Visualisation of the vector displacement

The displacement vectors were visualised as arrow glyphs at each voxel location using Paraview (http://www.paraview.org/), depicting the magnitude and direction of the displacement.

Each glyph was coloured according to the angle between the hypotenuse (resultant vector) and adjacent side in each plane (axial, sagittal and coronal). A circular colour map was created where the colour spectrum ranged from red-yellow-green-cyan-blue-red across a scale of $-\pi$ to π. All glyphs were then coloured according it angle (in radians) of displacement in each plane.

4.2.3.2.2 Brodmann area simulations examining the dependence of a putative function of interest on a single cluster of voxels

The relationship between damage and loss of function is likely to be far more complex than the simple voxel model described above. A more plausible alternative is the situation where damage to any part of a critical region can result in dysfunction of the whole region. It is reasonable to assume that the data collected in the majority of studies will have had a degree of noise interfering with the correlation between structure and function. We therefore introduced 10% noise into the outcome data to simulate this situation.

In this case we use the clusters defined by the standard Brodmann areas. Since we are not interested in the locus of a specific function but the level of bias present in the method, the arbitrary choice of Brodmann areas as the method of subdivision seems as good as any.

4.2.3.2.2.1 Statistical analysis

The boundaries for the Brodmann areas in MNI space were obtained from the Brodmann map found in MRIcro (Rorden and Brett, 2001). In order to describe

the vector of displacement, the centre of mass for each Brodmann area in the right hemisphere was calculated and treated as the ground truth locus, i.e. the locus associated with a specific brain function. A total of 41 Brodmann areas were examined.

The dataset of 581 binary masks were then classified into affected and unaffected groups depending on whether the lesion involved a minimum proportion of the Brodmann area across the percentage range between 5% and 60% in 5% steps. 10% noise was introduced by reversing the outcome label for 10% of the data, thus for 58 randomly selected subjects, a label of 0 was changed to1 and vice versa.

Similar to the voxel analysis, a 2x2 confusion table was set up at each voxel location to create a probability map for each Brodmann area simulation, with the frequencies derived from the 581 binary masks.

4.2.3.2.2.2 Calculation of the vector displacement

The probability map associated with each Brodmann area was thresholded at 1% Bonferroni correction for multiple comparisons, with the resultant cluster of voxels reduced to a single point locus by calculating the centre of mass.

A displacement vector was generated for each Brodmann area, describing the direction and magnitude of displacement from the “ground truth” Brodmann area centre of mass to the centre of mass of the locus identified by the mass univariate technique.

4.3 Results

4.3.1 Visualisation of a high dimensional data set in two dimensions using Isomap and tSNE

Figures 4.5 and 4.6 are the same 2 dimensional representation of the high dimensional relationship between the different lesioned brain masks. Each point within the 2 dimensional space represents a unique brain mask, where the distance between a pair of points is inversely proportional to their similarity. The algorithm attempts to arrange all the points within a reduced feature space, (105266 down to 2), while maintaining the relationships between them – an embedding. The extent to which the low dimensional embedding conserves the high dimensional similarities between the images, the distances from one point to another will reflect the similarity of that case to the other. If the differences between cases are merely noise, then the distances should be random, producing a random "starfield" where the points are evenly distributed. If, on the other hand, the differences between cases follow a pattern, with some cases being more similar to each in a consistent way, then the overall distribution will not be an evenly random field but a inhomogeneous field where the points fall into distinct clusters. This is what we observe here.

Figure 4.5 illustrates a centrifugal lesion volume gradient. More centrally located brains have smaller lesion volumes (white shaded areas), with lesion volume increasing as you move towards the peripheries (increasing blue shading). The change in lesion volume is not uniform across the 2 dimensional feature space, with a steeper rate of change along the horizontal axis, compared with the vertical.

Figure 4.6 highlights a distribution pattern in the 2 dimensional feature space that is determined by lesion location. There is a coarse separation of brains

according to lobe involvement illustrated by the different colours, however the transitions from one lobe to the next is not discrete.

The shading in figures 4.5 and 4.6 illustrates only 2 broad patterns identified by the tSNE algorithm. There are, however, further patterns present represented by the sub-clustering of points. The feature(s) tSNE is using to determine this sub-clustering is not clear, however, it does suggest that there are further patterns contained within the interaction between voxels leading towards this structured embedding.

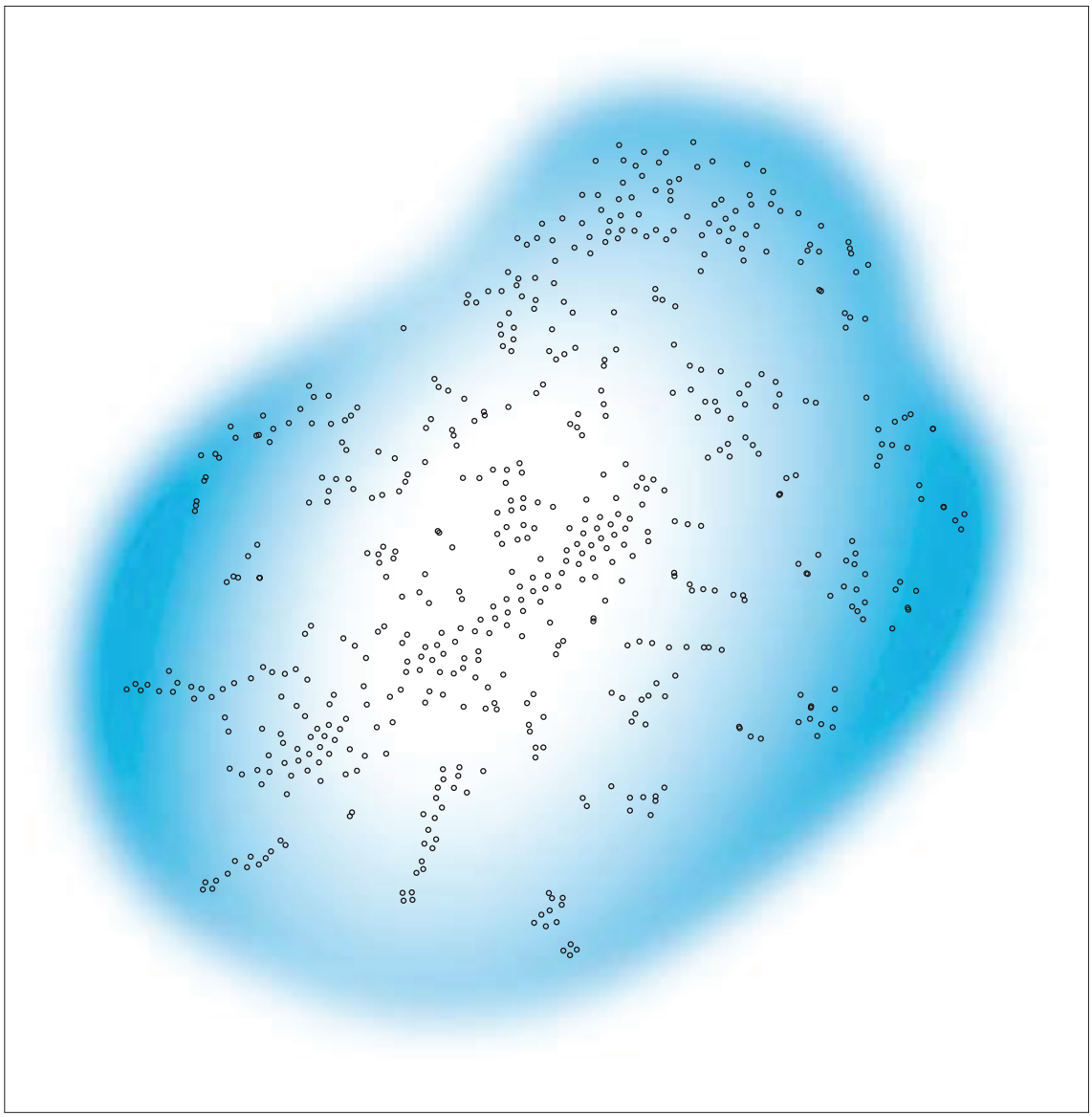


Figure 4.5 - A 2 dimensional embedding of the high dimensional data set of 581 brains using tSNE (lesion volume).

The high dimensional (105266 dimensions) dataset of 581 lesion brains has been embedded into 2 dimensions whilst trying to preserve the high dimensional interactions. Each voxel is considered as a separate dimension, with the stack of 581 lesions being first passed to Isomap to be reduced down to 30 dimensions. Then the 30 dimensions are reduced to 2 dimensions by using tSNE, with the resultant 2 values used as coordinates within a 2 dimensional space. Each lesion mask is represented by a point and there is pronounced clustering of the data suggesting that the voxels are not spatially independent. A blue colour gradient has been applied to provide an illustrative impression of the pattern of volume change across the 2 dimensional space from small (white) to large (blue).

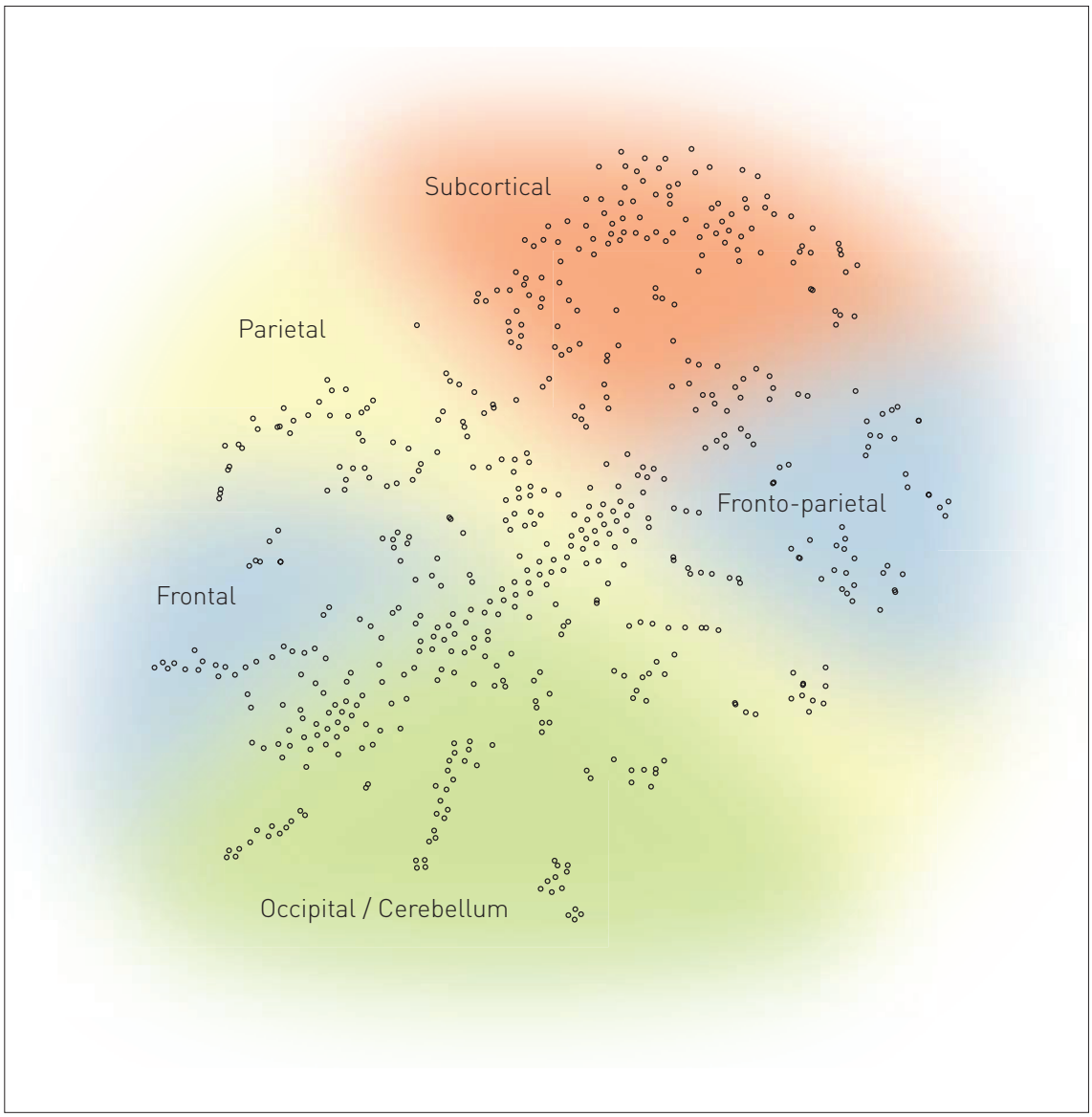


Figure 4.6 - A 2 dimensional embedding of the high dimensional data set of 581 brains using tSNE (lesion location).

The high dimensional (105266 dimensions) dataset of 581 lesion brains has been embedded into 2 dimensions whilst trying to preserve the high dimensional interactions. Each lesion mask is represented by a point and there is pronounced clustering of the data suggesting that the voxels are not spatially independent. The masks have been roughly distributed in the 2 dimensional space according to lobe involvement.

4.3.2 Mass univariate simulations – Dependence of a putative function of interest on a single voxel

The extent of mislocalisation across the brain is visualized as a vector field describing the magnitude and direction of the displacement from the true location at each voxel. Each vector is derived by thresholding the voxel's corresponding brain volume that represents the probability of associated involvement, at 1% significance, corrected for multiple comparisons with a Bonferroni correction.

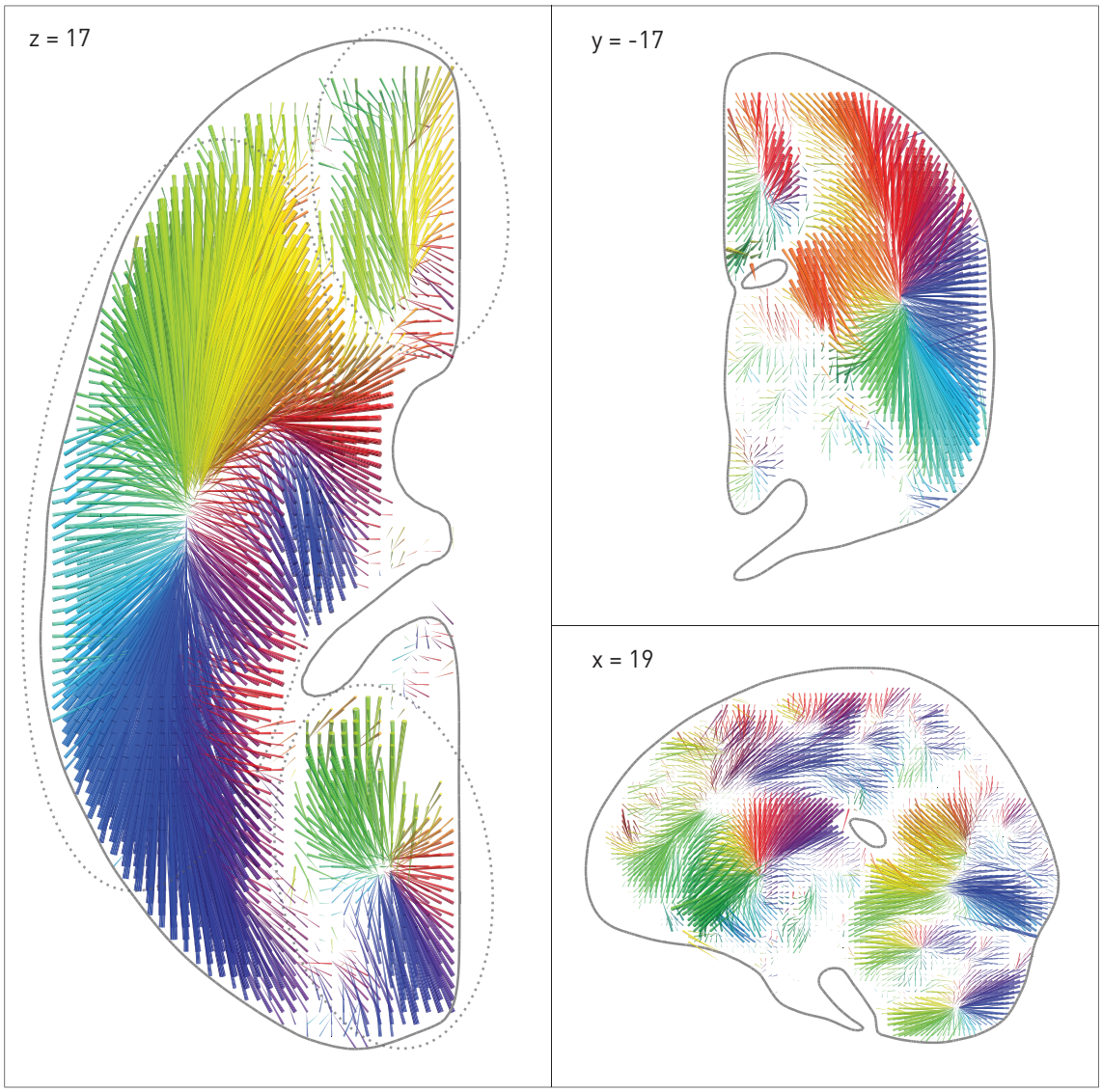


Figure 4.7 - Displacement vector map for the simulation using a single voxel dependence of a putative function of interest.

Three dimensional vector plot of the direction (colourmap) and magnitude (length of arrow) of mislocalisation at adequately sampled voxels within 3 representative planes (left axial z=17, top coronal y=-17, bottom sagittal x=19), based on a sample of 581 acute stroke lesions, normalised into standard stereotactic space and mirrored onto one hemisphere. The value at each voxel was calculated by labelling the stack of 581 lesioned volumes as being “affected” or “unaffected” depending on whether or not that voxel fell within the lesion in each volume, running a standard voxel-wise Fisher’s exact test-based mass-univariate analysis on the two groups, and identifying the centre of the resultant significant cluster, identified by the asymptotic *p* value thresholded at a Bonferroni corrected $p<0.01$. This procedure was performed at all voxels hit more than 3 times in the dataset. Note that the mislocalisation tends to follow the organisation of the vascular tree, with clusters corresponding to the branches of the middle cerebral, anterior cerebral, and posterior circulations (dotted rings). The mean displacement is 15.7mm (standard deviation 9.2mm, median 14.1mm, interquartile range 11.8mm).

All data was collapsed onto one hemisphere. The voxel locations which do not have an associated glyph vector are relatively few and are secondary to fewer than 4 lesions involving it.

Substantial error is observed in all regions of the brain, with a mean of 7.8 voxels (15.7mm) and standard deviation of 4.6 voxels (9.2mm). The direction of the displacement is not random, with a centripetal movement from cortical loci to deep white matter. Figure 4.7, displays a 2 dimensional slice through the hemisphere in each of the 3 planes where the glyph vectors are coloured according to the direction of displacement within the 2 dimensional plane (2π radians, axial (z=17), coronal (y=-17) and sagittal (x=19)).

In the axial plane, there are 3 distinct clusters visible, illustrated in figure 4.7 (dotted rings). The largest of these depict a shift towards the vicinity of the middle cerebral artery. The other 2 clusters are located at the frontal and occipital poles, with the former displaying vector displacements towards the anterior cerebral artery. The caudal-rostral extent of the occipital cluster can be clearly visualised in the coronal image slice.

4.3.3 Mass univariate simulations – Dependence of a putative function of interest on a single Brodmann area

In the single Brodmann area analysis the hypothetical critical regions were not individual voxels but groups of voxels falling into a standard Brodmann area. Figure 8.2 in appendix A is a table listing the mislocalisation incurred for different Brodmann areas when a lesion mask was considered “affected” at different percentage involvements. In the scenario where a minimum percentage involvement was 20% the mean error was 8.0 voxels (15.9mm) with standard deviation of 8.8 voxels (17.6mm).

	%											
	5	10	15	20	25	30	35	40	45	50	55	60
mean	18.60	17.34	14.78	15.92	14.46	16.22	13.13	11.75	10.24	10.70	8.84	10.96
std	7.28	8.74	7.51	17.61	7.22	8.16	6.82	4.90	5.65	6.21	3.94	4.98
iqr	11.66	11.67	9.50	9.00	10.68	9.22	8.98	5.77	6.80	10.22	2.47	4.82

Figure 4.8 - Table of the displacement means, standard deviations and inter-quartile ranges for the 41 assessed Brodmann areas.

The displacement means, standard deviations and inter-quartile ranges (mm). The minimum percentage volume involvement of a Brodmann area that determined whether a lesion volume was “affected” or “unaffected” was increased in 5% steps. A standard voxel-wise Fisher’s exact test-based mass-univariate analysis was then performed on the two groups, and the centre of the resultant significant cluster identified by the asymptotic p value thresholded at a Bonferroni corrected $p<0.01$.

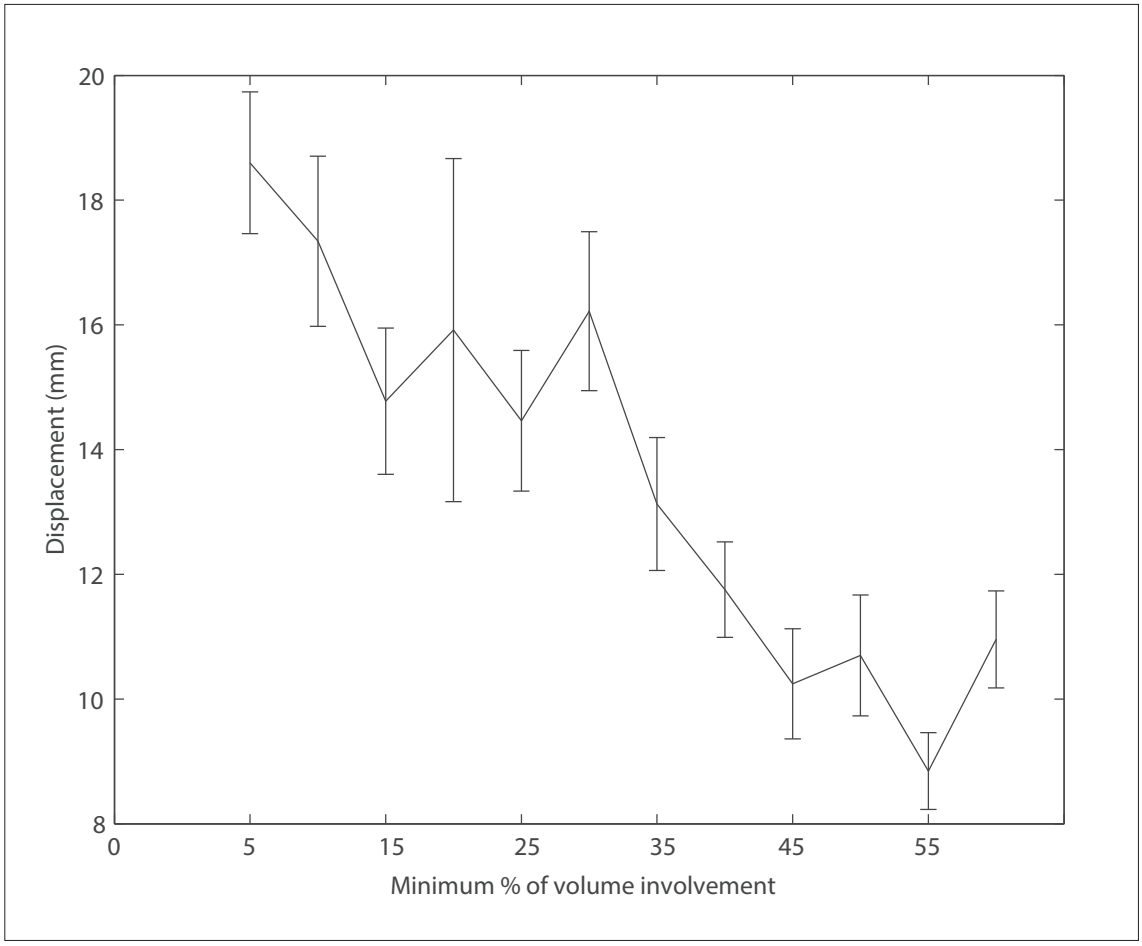


Figure 4.9 - Plot of the mean displacement for 41 Brodmann areas as a function of minimum percentage volume involvement.

The mean mislocalisation of the 41 assessed Brodmann areas as a function of the minimum percentage volume involvement necessary to label a lesion volume as “affected”. The error bars are the standard error for each mean.

4.4 Discussion

The foregoing simulations demonstrate that there is a hidden structure that can lead to bias within vascular lesion data. Both Isomap and tSNE are dimensionality reduction algorithms that attempt to cluster the binary lesion masks into a 2 dimensional space whilst trying to preserve the high dimensional relationships between them.

If the distribution had been completely homogeneous, then the 2 dimensional embeddings would have been homogenous too. This is however not the case, as illustrated in figures 4.5 and 4.6, where the embeddings show distinct clusters rather than a uniform distribution of points. At the gross level the brains have been organised according to both lesion volume and location (figures 4.5 and 4.6). Although it is not possible to identify all the features tSNE uses to determine its 2 dimensional embedding, these two patterns would suggest that the distribution of points within this 2 dimensional space is not entirely random. The presence of clusters within the embedding would suggest that there are further patterns within the data which are not as easily described by this visual representation. The number and type of patterns that actually exist is not clear, however, it would be reasonable to state there is stereotypy within the data that is a potential source of bias.

This is confirmed by the single voxel simulations where the displacement vectors, representing the mislocalisation introduced by assuming spatial independence, are created for each voxel within the brain. Distinct clusters are formed in the frontal, occipital and temporo-parietal regions (figure 4.7). These clusters demonstrate a shift from cortical to subcortical structures that appear to follow the underlying vascular tree. Interestingly the boundaries defined by the clusters are reflective of the watershed areas between the different cerebral arteries (Hendrikse et al., 2008; van Laar et al., 2008; Savoiardo,

1986). Therefore by ignoring the spatial relationships between voxels – which is exactly what is done when we treat each voxel as spatially independent in a mass univariate approach – we will inevitably incur a mislocalisation.

Irrespective of the influence of each of these biases, a mass univariate approach leads to a mean mislocalisation of 15.7mm, a distance sufficient to move across lobes, Brodmann areas and from grey to white matter. This may in fact be a rather conservative estimate, as the Bonferroni correction tries to minimize false positive occurrences, at the cost of producing false negatives. In other words the Bonferroni correction will reduce the sensitivity of the mass univariate technique of finding a significant region and ultimately shrink the volume that falls within a *p* value of 1% corrected for multiple comparisons. If the relation between damage and loss of function is more complex than the simple isolated voxel model described, the mislocalisation can only be greater. Specifically, if damage to any part of a critical region can result in dysfunction of the whole region the maximally significant locus identified by conventional lesion-mapping need not even include any part of the critical region.

To simulate this more physiologically plausible model the analysis was repeated with the hypothetically critical regions being not individual voxels but groups of voxels falling into standard Brodmann areas. The mislocalisation of each Brodmann area independently was evaluated, labelling an area as inactivated in a given brain if more than 20% of its constituent voxels were affected. There is a mean error of 15.9mm with a standard deviation of 17.6mm, larger than the simpler, and less physiologically plausible analysis.

Quantifying the error in this way may be thought to provide the means of eliminating it. Unfortunately the inverse of the error vector field is only valid for the underlying lesion-function model, the one entity we are trying to identify. Consequently unless we know that the functions we wish to examine are related to single foci in the brain, these vector maps cannot be used as

a corrective solution. Considering the highly distributed organisation of the brain, we must therefore determine the error with models that are dependent on multiple loci within the brain. Although the number of loci and possible combinations are exhaustive, it is reasonable to extrapolate that if the mass univariate approach is susceptible to systematic bias in a two loci model, arguably the simplest multiple loci model, this will continue to be the case as the number of loci increase. If the latter is demonstrated then in no scenario will the mass univariate approach be insusceptible to the mislocalisations caused by these latent systematic biases within the data.

4.5 Conclusions

Vascular lesion data has, for the past few decades, been the main type of data used in lesion function mapping studies, to try and decipher the underlying relationship between structure and function. The following simulations have quantified the amount of mislocalisation that casts a shadow of doubt over inferences made from previous lesion-function mapping studies. Importantly, quantifying the error in this way cannot be used to eliminate it as the inverse of the error vector field is only valid if the underlying lesion-function model is valid. Given the fundamentally distributed organisation of the brain, a model consisting of multiple loci must be considered.

5 Lesion-function inference in the context of spatially distributed function

5.1 Introduction

There have been growing concerns over the biases present within vascular lesion data and their effect on the inferences obtained from lesion-symptom mapping studies particularly with techniques such as VLSM (Nachev, 2008). In the previous chapter, the magnitude of mislocalisation encountered with a mass univariate technique was quantified for the single locus scenario across the whole brain. By assuming spatial independence between voxels, the models simulating the dependence of a putative function on a single voxel locus, produced a mean displacement of 15.7mm from the true locus to the one identified by the technique. Although connectivity in the brain is dominated by local connections it is clear that areas that are anatomically remote may nonetheless be functionally related (Barbas and Mesulam, 1981; Bates and Goldman-Rakic, 1993; Callaway, 1998; Catani and Thiebaut de Schotten, 2008; Catani et al., 2012; Dougherty et al., 2005; Katz et al., 1989; Leh et al., 2007; Petrides and Pandya, 2002; Stanton et al., 1995; Thiebaut de Schotten et al., 2011). For this reason the relation between damage and loss of function is unlikely to be as simple as these single locus models that ignore these long range connections (Catani et al., 2002; Mesulam, 1981). Instead more complex models that involve multiple loci are more likely to exist.

If we take for example the symptom of spatial neglect, single case studies have shown that damage to either the frontal or parietal lobe can precipitate the functional deficit (Damasio et al., 1980; Husain and Kennard, 1996; Mort et al.,

2003; Vallar and Perani, 1986). Moreover, this relationship between anatomy and function is not linear with a case of severe neglect following a right parietal lesion being corrected after a second left sided frontal stroke (Vuilleumier et al., 1996).

These case studies demonstrate that not only can multiple areas precipitate the same functional outcome, but also the interaction between these loci can influence the outcome. Therefore, in order to appropriately model the brain, we should account for multiple loci and the interactions between them. More recently Karnath et al (2001) applied a mass univariate technique to a cohort of patients with and without neglect, where each voxel within the brain was treated as spatially independent. Contrary to the previous case studies, they reported the locus of neglect was a function of the temporal not the parietal lobe (Karnath et al., 2001). The differences in localisation between the case studies and group studies raises concerns over whether the spatial inferences obtained from mass univariate analyses are indeed true. Importantly if multiple loci models exist within the brain, can we proceed with a mass univariate technique with any confidence?

One potential solution would be to generate models using high-dimensional inference that try to capture the multivariate lesion distribution. Rather than treating each location in the brain in isolation, we model the pattern of damage across the entire brain, inevitably using thousands of variables. In order to assess the multivariate distribution across the brain volume, the data supplied to the algorithm must be parameterised in a spatial manner, in this case with each voxel location treated as a separate dimension. The algorithm must then assess the interactions that exist between each and every voxel and all the possible combinations. Estimating such models requires non-traditional inferential methods based on machine learning. These models are essentially

classifiers, and in the case of lesion symptom mapping, are aimed at determining which regions of the brain are correlated with a specific outcome.

Smith et al (2013) used the multivariate technique of Support Vector Machines (SVM) in a lesion mapping study of spatial neglect. They demonstrated the predictive performance of the multivariate technique was better than their univaritate approach and that the superior temporal gyrus contained significant information to facilitate the classification of neglect and control subjects. However, in their later analyses they reduced the dimensionality of the data from the whole brain feature space down to pre-specified regions of interest (ROI). First they examined each ROI *independently* from the other, though a multivariate technique was applied to the voxels within the ROI in question. Second, 12 perisylvian ROIs were selected, from which combinations of 2 or 3 ROIs were used. Although this increased the dimensionality of the feature space, the 12 ROIs were chosen based on previous studies using a mass *univariate* technique. As a result this ignored the influence of voxels, similarly sampled by the lesion dataset, filtered out because of a technique that previously assumed independence across voxels.

The process of assessing the inter-dimensional interactions raises a number of issues. First, as the number of dimensions increase linearly, the total number of inter-dimensional interactions increases according to a power rule. As a consequence large data sets are necessary to ensure these relationships are thoroughly interrogated to avoid over-fitting. Conceptually, if there is a mismatch where the total number of dimensions exceeds the total number of data points (dimensionality mismatch), then some degree of over-fitting will occur secondary to the sparsity of the data within the feature space. In our case, the total number of dimensions will exceed a thousand and may be in the order of tens of thousands. Although theoretically such large datasets are possible, pragmatically we are considered fortunate to have samples in

the hundreds. Consequently the curse of dimensionality remains a concern for all algorithms. The second issue is computation time. Although there are techniques to minimise the number of calculations, simplistically each inter-dimension interaction must be modelled, therefore as the number of dimensions increase so will the computational load.

There are a number of techniques currently available that can be used to model high dimensional datasets. One such technique is Support Vector Machines (SVM). Developed in the 1990s, SVM attempts to find a dividing hyperplane that accurately separates the two outcome classes associated (§1.3.3.2.1), labelling the data points according to their location either side of the hyperplane (Boser et al., 1992; Cortes and Vapnik, 1995; Vapnik, 1982). The popularity of SVM has increased recently, as they have been found to cope relatively well with the dimensionality mismatch problem (Saur et al., 2010). The SVM model is defined by a sub selection of data points, known as support vectors (SVs), which help describe the position of the hyperplane within the high dimensional feature space. Although it is a sparse model, the total number of SVs will grow linearly with the size of the training set. Consequently increasing the training data will inevitably increase the computational load (Chen et al., 2009).

Alternatively a Bayesian approach may be used. In this case the dataset is used to derive the posterior probability, which is essentially the likelihood a new data point will belong to a specific class. This posterior probability is encoded in the model and reflected as the combined relevance or weights of each dimension. Iterative processes, such as Markov Chain Monte Carlo Bayesian inference (§1.5.2.4), have been developed that facilitate repeated sampling from the posterior probability that can eventually tend towards the true posterior (Neal, 1993). Understandably in a full Bayesian model, where each dimension is attributed a weight, this can be extremely computationally

intensive. One solution to this problem is to use a sparse Bayesian technique where dimensions deemed unimportant are excluded from the model and relevant ones are preserved in an iterative process. This is the basis of the Relevance Vector Machine (RVM), which like SVM, produces a sparse model (Tipping, 2001). In contrast to SVM, the Bayesian framework of RVM means that a grid search of parameters is not required, therefore simplifying the process of identifying the optimal model.

The following chapter will assess the impact of a distributed neural dependence where a deficit will occur when either of the two areas reaches a minimum volume of damage. The combinatorial expansion makes comprehensive modelling of this computationally prohibitive; nonetheless if a problem is shown for one biologically plausible pairing then no hypothetical pairing can be trusted. The pairing of Brodmann areas 39 and 44 will be examined, as two brain regions which, when lesioned, can independently lead to the syndrome of neglect (Damasio et al., 1980; Husain and Kennard, 1996; Mort et al., 2003; Vallar and Perani, 1986). Two series of simulations will be performed, the first will assess the performance of a mass univariate technique against the multivariate technique of SVM. The second series of simulations will then compare the performance of 3 different multivariate methods; Support Vector Machines; a sparse Bayesian technique, Relevance Vector Machines; and a full Bayesian approach, Flexible Bayesian Modelling by examining their ability to make spatial inferences and the predictive accuracy of the model on a separate test dataset.

5.2 Methods

5.2.1 Imaging

A total of 581 brain scans were collected from the National Hospital for Neurology and Neurosurgery for this series of simulations. All brain scans possessed a vascular lesion and were performed on a GE Genesis Signa 1.5 Tesla (1.5T) MRI scanner in a single session for each patient. The scans collected were the axially acquired T2-diffusion weighted echoplanar images (b0 and b1000 sequences) sampled at 1mm x 1mm x 6.5mm (TR=10000ms, TE=104.9ms, Columns=256, Rows=256, Slice thickness=5mm). The mean age was 62.3 years (standard deviation (SD) = 17.8 years), and the proportion of males was 0.547.

Note that the simulations and analyses that follow do not depend upon any knowledge of whether any of these patients suffered from spatial neglect. Rather, the aim here is to determine lesion localisation for a syndrome that arises, in theory, from damage independently either to voxels in Brodmann's area 39 or 44.

5.2.1.1 Image preprocessing

For each patient the corresponding b0 and b1000 images were reconstructed into a 3 dimensional volume using xmedcon (Nolf et al., 2003). The b1000 volume was rigidly co-registered to the b0 volume using the co-registration function in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and its default parameter settings (the specific parameter settings are listed in appendix B). The b0 volume was subsequently spatially normalised into MNI space (ICBM space – European brains – template) using the combined segmentation-normalisation method in SPM8 (Crinion et al., 2007), with the derived normalisation

parameters applied to the co-registered b1000 sequence. Again the default parameter settings for the segmentation-normalisation method provided by SPM were used, except for the interpolation setting in the writing option being set to a 6th degree spline.

A binary mask for each patient was created via the lesion segmentation using zeta anomaly detection method described in §2 (Mah et al., 2012). The resultant masks were sliced at 2mm isovoxels and occupied a volume measuring 91 x 109 x 91 voxels. Figure 5.1 displays the volume coverage of the 581 different lesions.

Although there is evidence to show an increased preponderance to injury to the left hemisphere secondary to the anatomical arrangement of the aortic arch, it is generally believed the distribution within each hemisphere is similar (Naess et al., 2006). Consequently to increase the density of lesion coverage per voxel, all unilateral left sided lesions were reflected in the mid-sagittal plane. For bi-hemispheric lesions, the hemisphere with the larger lesion load was treated as the dominant side, with left hemisphere dominant bi-hemispheric lesions reflected in the mid-sagittal plane. Figure 5.2 displays the volume coverage of the right hemisphere after collapsing the binary masks onto a single hemisphere.

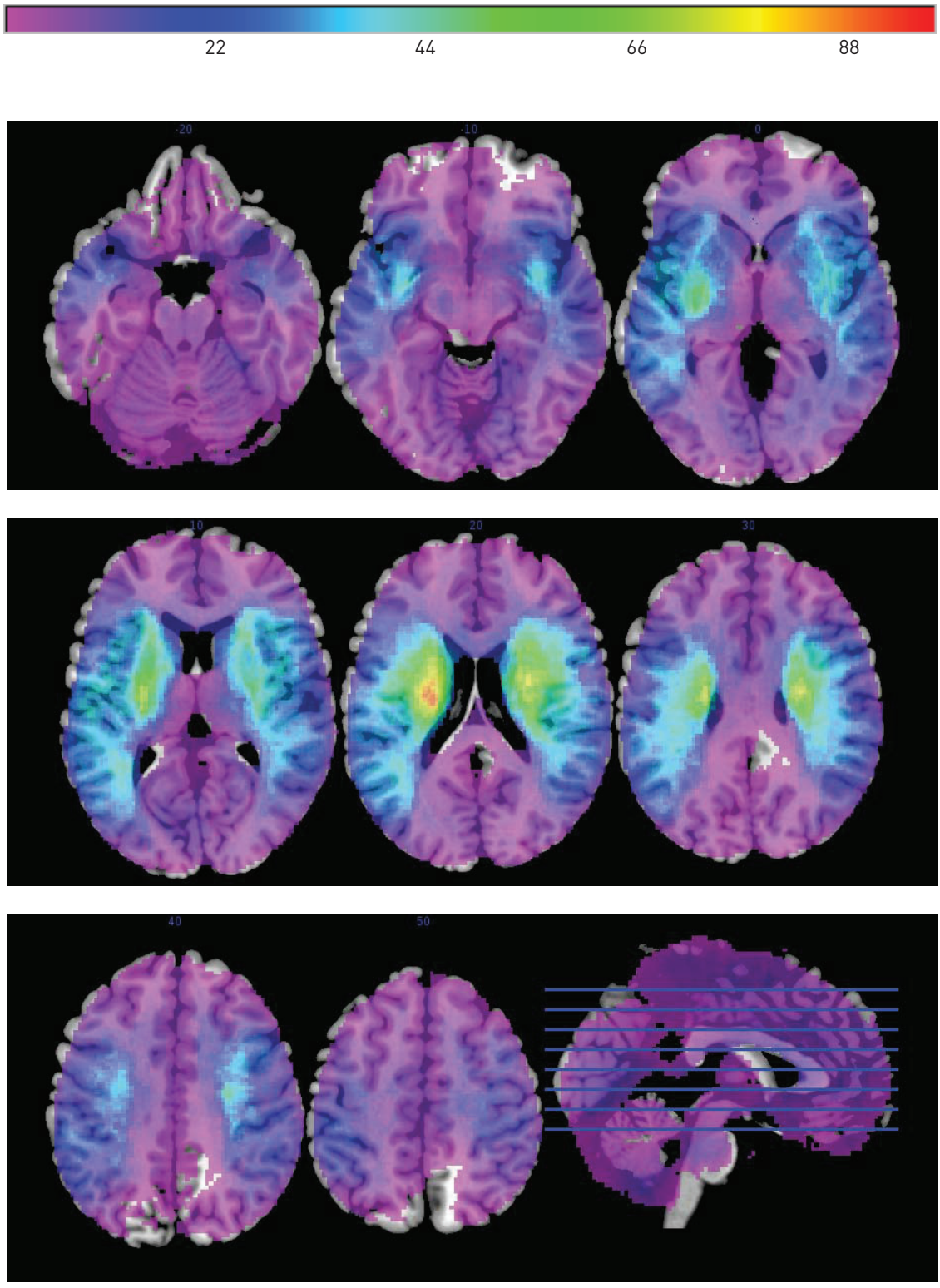


Figure 5.1 - Lesion overlay map of the 581 lesion masks.

The axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. The peak overlay value is 96. The distribution of lesion injury is symmetrical across the mid-sagittal plane.

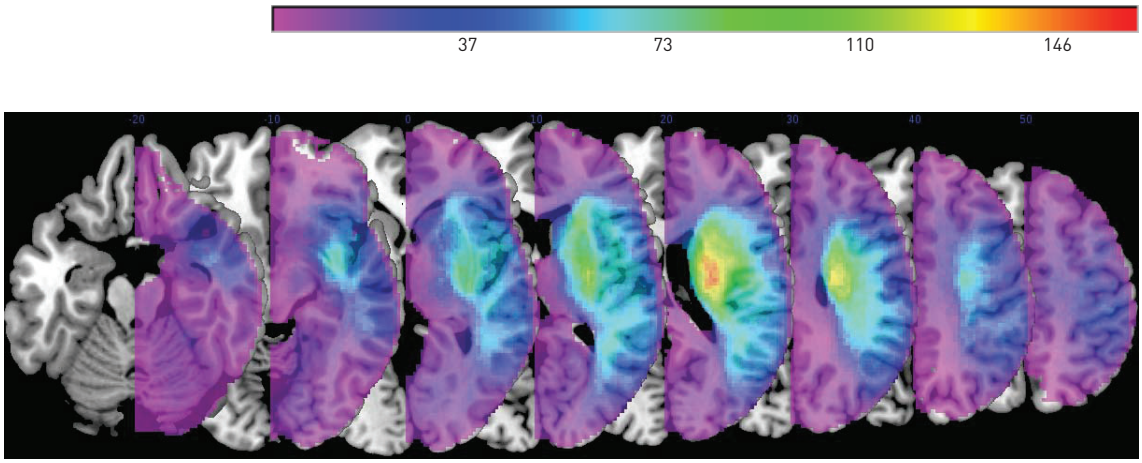


Figure 5.2 - Lesion overlay map of the 581 lesion masks collapsed on the right hemisphere.

The axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. The peak overlay value is 160.

5.2.2 Hardware

All simulations were performed in the MATLAB environment (The MathsWorks Inc.). A dual processor Intel Xeon 5600 2.4GHz processor, with 24GB RAM and a 120GB solid state drive (SSD) was used for both the mass univariate and Support Vector Machine models in simulation 1.

In simulation 2, the model created by the Flexible Bayesian Modelling technique was performed on a dual processor Intel Xeon 5600 2.4GHz processor, with 24GB RAM and a 120GB solid state drive (SSD), while the Support Vector Machine and Relevance Vector Machine models were performed on an Intel i7 920 2.66GHz processor, with 6GB RAM and 120GB solid state drive (SSD).

5.2.3 Simulations

For the subsequent simulations only the within brain voxels were used. The brain mask template from SPM8 was thresholded at less than 0.8, and greater than 0.4 to identify the within brain voxels. Since the lesions have been collapsed onto the right hemisphere, only those voxels in the right hemisphere were then identified and used to create a 581 by 105266 data matrix.

As mentioned earlier a comprehensive evaluation would be unhelpful as we do not know that the Brodmann parcellation, or any other parcellation, is truly representative of the underlying functional anatomy. The only necessity is a hypothetical rule which we can use as the “ground truth” against which to compare the performance of the different models. Brodmann areas 39 and 44 were chosen (two putative loci for visio-spatial neglect) as the 2 discrete and spatially separate clusters, identified using the template provided in the MRIcro suite (Rorden and Brett, 2001).

5.2.3.1 Simulation one : Comparison of a mass univariate (Fisher's exact test) technique against a multivariate (SVM) technique

5.2.3.1.1 Lesion symptom model

The hypothetical model used in this series of simulations was the dependence of a putative function to the injury of 20% or more of either Brodmann area 39 (BA39) or 44 (BA44) resulting in a hypothetical deficit of interest 90% of the time. In other words, there is a 90% chance that a lesion affecting 55% of BA 44 and 5% of BA 39 would result in expression of the symptom, while 15% involvement of BA44 and 15% involvement of BA 39 would not.

5.2.3.1.2 Data preparation

The stack of 581 binary lesion masks was used at a resolution of 2mm isovoxels. Each image was classified as "affected" or "unaffected" depending on whether 20% or more of either BA39 or BA44 fell within the lesion present in the scan. This resulted in a split of 98 affected and 483 unaffected masks. Similar to the method in the previous chapter, 10% noise was introduced into the dataset. Therefore for 58 randomly selected subjects, a label of 0 was changed to1 and vice versa. This generated a dataset with a split of 445 unaffected vs. 136 affected.

5.2.3.1.3 Mass univariate analysis (Fisher's exact test)

Only voxels that had more than 3 binary masks affecting it were included in the analysis. For each of the 90469 voxel locations hit at least 4 times in the dataset a 2 x 2 confusion table of "hit" or "miss" and "affected" or "unaffected" was constructed. The frequencies for each table were derived from the dataset of 581 binary masks classified by the above hypothetical rule. This produced

a *p* value representing the likelihood a voxel within the brain will by chance be injured if the scan is associated with the hypothetical functional deficit. Voxels with smaller values are more likely to be correlated with expression of the symptom.

The *p* values derived from the voxelwise Fisher's exact test performed at each of the 90469 voxels were mapped into its corresponding location within the brain volume (Fisher, 1970). This produced a probability map of the brain indicating the locations whose injuries are significantly correlated to the presence of the functional deficit.

5.2.3.1.4 Multivariate analysis (SVM)

There is a total of 105266 voxels within the right hemisphere. Only voxels that were involved by 4 or more of the 581 lesion masks were passed to the algorithm. This resulted in a dimensionality mismatch of 90469 dimensions (voxels) to 581 data points.

The program LibSVM (Chang and Lin, 2001) was used to generate the model. In order to identify the optimal model using Support Vector Machines, a parameter search for the *C* value in the linear kernel was performed, using a 17 fold cross-validation technique. A total of 41 different *C* values were assessed, where $C = 2^x$, and $-20 \leq x \leq 20$ for all integers of *x*. To compensate for the disparity in sample sizes between the affected and unaffected groups the affected group was given a weighting of 3. The *C* value which generated the model with the highest level of accuracy from the 17 fold cross-validation was subsequently used to generate the optimal linear model using the entire 581 data set.

The weight associated with each dimension – in this case each voxel – was extracted from the optimal model. The weights have 2 properties; polarity,

indicating which class, affected or unaffected, the dimension is predictive, and magnitude which provides a guide to the relative importance of the voxel with respects to the other voxels within the brain. To evaluate the method's ability to identify the pre-specified critical regions, the extracted weights generated by the linear kernel were mapped into their corresponding locations within the 3 dimensional brain volume.

5.2.3.1.5 Comparison of the mass univariate with the multivariate technique

The mass univariate technique that uses a Fisher's exact test generates a *p* value for each voxel describing the likelihood the voxel will also be "hit" given the presence of the symptom, while the Support Vector Machine model produces a weight for each voxel, that indexes its contribution to the classification process. These 2 measures are not directly comparable and therefore to compare their performance a common feature must be identified between them. In this case the total number of voxels within BA44 and BA39 were identified (1705+1061=2766) and treated as 100%. Next stepwise percentages of this volume were selected with the corresponding threshold values necessary to isolate the same number of voxels in either the SVM weighted brain map or the Fisher's exact probabilities brain map. Theoretically, if the fidelity of the model to the hypothetical rule was perfect, the threshold at 100% would generate an image where all voxels within BA44 and BA39 are identified. The performance of both techniques could then be visually compared. Also since we know the locations of the voxels that constitute the hypothetical rule, the accuracy of the models along the range of percentages from 0-100% can be calculated, and a receiver operating curve (ROC) can be plotted.

5.2.3.2 Simulation two : Comparison of different multivariate techniques

In the second series of simulations, three different multivariate techniques were compared against each other: Support Vector Machines (SVM), Relevance Vector Machines (RVM), and Flexible Bayesian Modelling (FBM). The model generated by each technique was assessed for its ability to (i) identify the critical regions determining the presence or absence of the symptom as defined by the hypothetical rule, and (ii) predicting whether a novel brain volume was affected or unaffected by the hypothetical functional deficit based solely on the spatial parameters of the lesion.

5.2.3.2.1 Lesion symptom model

The hypothetical model used in this series of simulations was the dependence of a putative function to the injury of 30% or more of either Broadmann area 39 (BA39) or 44 (BA44) resulting in a hypothetical deficit of interest. In other words, a lesion affecting 55% of BA44 and 5% of BA39 would result in expression of the symptom, while 20% involvement of BA44 and 20% involvement of BA39 would not.

5.2.3.2.2 Data preparation

To reduce the computational complexity, the stack of 581 binary lesion masks was converted to a resolution of 4mm isovoxels. This was achieved by reslicing each binary mask via the SPM8 software suite and a modified re-orient function (appendix B).

The stack of resliced binary masks were then classified as “affected” or “unaffected” depending on whether 30% or more of either BA39 or BA44 fell within the lesion present in the scan. There is a total of 12938 voxels within

the right hemisphere after reslicing. Only voxels implicated in 4 or more binary masks were used in the subsequent simulations resulting in a total of 9638 voxels.

To assess the predictive performance of each technique the dataset of 581 scans was randomly divided into a training and test set. The training set consisted of 425 images with an “affected” and “unaffected” split of 52 and 373 respectively. All three techniques were kept naive to the test set during training.

The test set consisted of 156 scans, with an “affected” and “unaffected split of 21 and 135. The test set was further stratified into a balanced subset of 21 “affected” scans and 21 randomly selected “unaffected” scans.

5.2.3.2.3 Support Vector Machines (SVM)

The program LibSVM (Chang and Lin, 2001) was used to generate the model. A parameter search for the *C* value in the linear kernel was performed, using a 17 fold cross-validation technique. A total of 41 different *C* values were assessed, where $C = 2^x$, where $-20 \leq x \leq 20$ for all integers of *x*. To compensate for the disparity in sample sizes between the affected and unaffected groups defined by the hypothetical rule, a weighting of 7 was applied to the affected group.

The *C* parameter that produced the highest accuracy from the 17 fold cross-validation training was then used to generate the optimal linear model derived from the entire 425 data set.

The weights associated with each voxel were mapped into its corresponding location within the brain volume.

5.2.3.2.4 Relevance Vector Machines (RVM)

The program Sparse Bayes (Tipping, 2001) was used to generate a predictive model using Relevance Vector Machines. The likelihood parameter was set to "Bernoulli", with the number of iterations restricted to a maximum of 5000. The remaining parameters were left as the default setting of the software.

Two forms of analysis were performed using RVM. The RVM model only provides weights for dimensions it identifies as relevant for outcome prediction, while the remaining dimensions are given a zero weighting. To try and overcome this issue a sub selection of 423 scans from the training dataset was selected and an RVM model created. This process was iterated 212 times to create 212 different models. The relevance vectors (RVs) from the 212 different models were then summed across the brain volume, with those voxels not deemed an RV being given a weight of zero by the algorithm. The mean weight for each voxel was then calculated.

The second form of analysis was to facilitate comparison of the RVM prediction performance with the other multivariate approaches. In this simulation, the entire stack of 425 binary masks was used to derive the RVM model.

5.2.3.2.5 Flexible Bayesian Modelling (FBM)

The program flexible Bayesian modelling (FBM) was used to generate a predictive model using a Markov Chain Monte Carlo (MCMC) Bayesian inference approach (Neal, 2004). A hybrid Gibbs sampling method was chosen, with a chain length of 1000, step size of 0.05 and a decay of 0.2. The simulation was either iterated 1000 times with a burn-in period of 500; or iterated 2000 times and a burn in period of 1000.

The weights created for each voxel by FBM varies from iteration to iteration with the process approaching the true weight asymptotically. The weights which were mapped into their corresponding location within the brain volume for the purpose of visualisation were obtained by calculating the mean value of the weights from the latter half of the iterative process. For example, each voxel in the 1000 iterations simulation will have a set of 500 weights after removal of the burn-in period. The displayed weight will therefore be the mean of this set of 500 values.

5.2.3.2.6 Comparison of SVM, RVM and FBM

Both the SVM and FBM models provide a weight for each dimension (voxel), while the RVM model, due to its sparseness only provides a weight for the dimensions it deems “relevant”. In all three cases a different scale is used for the weights, therefore similar to simulation 1 the total number of voxels within BA44 and BA39 were identified (213+132=345) and treated as 100%. Next ascending in a stepwise manner, different percentages of this volume were selected by identifying the threshold values necessary to isolate the same number of voxels with the FBM and SVM models. For the RVM model this was not possible due to its sparsity. Instead all the voxels which were identified as positively predictive for the presence of the deficit (positive relevance vectors) were identified. The pattern of functional localisation by the RVM model was then assessed as a function of the percentage of total positive RVs.

Similar to simulation 1, the locations of the voxels that constitute the lesion symptom rule are known. Therefore the accuracy of the FBM and SVM models can be compared along the range of percentages from 0-100% and a receiver operating curve (ROC) can be plotted.

The predictive performance of the models was assessed by examining the sensitivity, specificity, positive predictive value and negative predictive value on the novel test data set of 156 brains. The performance was assessed on a balanced dataset of 42 “affected” and “unaffected” scans as well as the entire 156 images.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$positive\ predicitive\ value = \frac{TP}{TP + FP}$$

$$negative\ predictive\ value = \frac{TN}{TN + FN}$$

True positive (TP) = correctly identified as affected

True negative (TN) = correctly identified as unaffected

False positive (FP) = incorrectly identified as affected

False negative (FN) = incorrectly identified as unaffected

5.3 Results

5.3.1 Simulation 1 : Comparison of a mass univariate (Fisher's exact test) technique against a multivariate (SVM) technique

The mass univariate technique (Fisher's exact test) creates a probability map of the likelihood a voxel will be lesioned by chance given the presence of a lesion, whilst the multivariate approach of SVM generates weights assigned to each dimension (voxel) which indexes its contribution to the classification process. Both scales allow the comparison of the relative importance of a voxel to its neighbour, but not across techniques. To facilitate the comparison across techniques, a threshold value for each scale is identified where a set percentage of the total true critical region is identified (BA39 and BA44). The voxels remaining after thresholding can lie anywhere within the brain, and its location is dependent on how accurately the analysis technique models our hypothetical rule.

5.3.1.1 Mass univariate analysis (Fisher's exact test)

The mass univariate analysis using a Fisher's exact test took 3 minutes to generate the model.

The voxel with the lowest probability of concomitantly being lesioned by chance given the presence of the functional deficit, i.e. the peak, is located along the anterior, inferior aspect of Brodmann area 39 (BA39) (figure 5.3 – left column). As the number of inclusion voxels is increased to 5% of the total true volume (138 voxels) (figure 5.4 – left column), the peak voxel beings to form a cluster within BA39. A second smaller cluster appears in Brodmann area 44

(BA44) along the posterior boarder, with the space between the two critical regions populated by a few sparse voxels.

By 10% of the true volume (276 voxels) (figure 5.5 – left column), both clusters in BA39 and BA44 continue to expand in size, with some slight extension beyond the anterior, inferior border of BA39. There is an established cluster of voxels between BA39 and BA44, that inhabit the region of the superior temporal gyrus (STG). Importantly the centre of this significantly associated region for the mass univariate approach does not fall in either Brodmann area, but in the region of the superior temporal gyrus. Figures 5.3 through to 5.9 show the outline of BA44 and BA39 in grey, with the thresholded voxels from the mass univariate technique in red. The location of the STG is outline in blue. The level of the brain outline in the axial perspective is z=17, while the coronal view is at y=-17.

Further expansion of the 3 clusters continues, when at 20% of the true volume (553 voxels) the cluster in the STG merges with the cluster in BA39 (figure 5.8 – left column). The separation between the STG cluster and BA44 cluster is maintained.

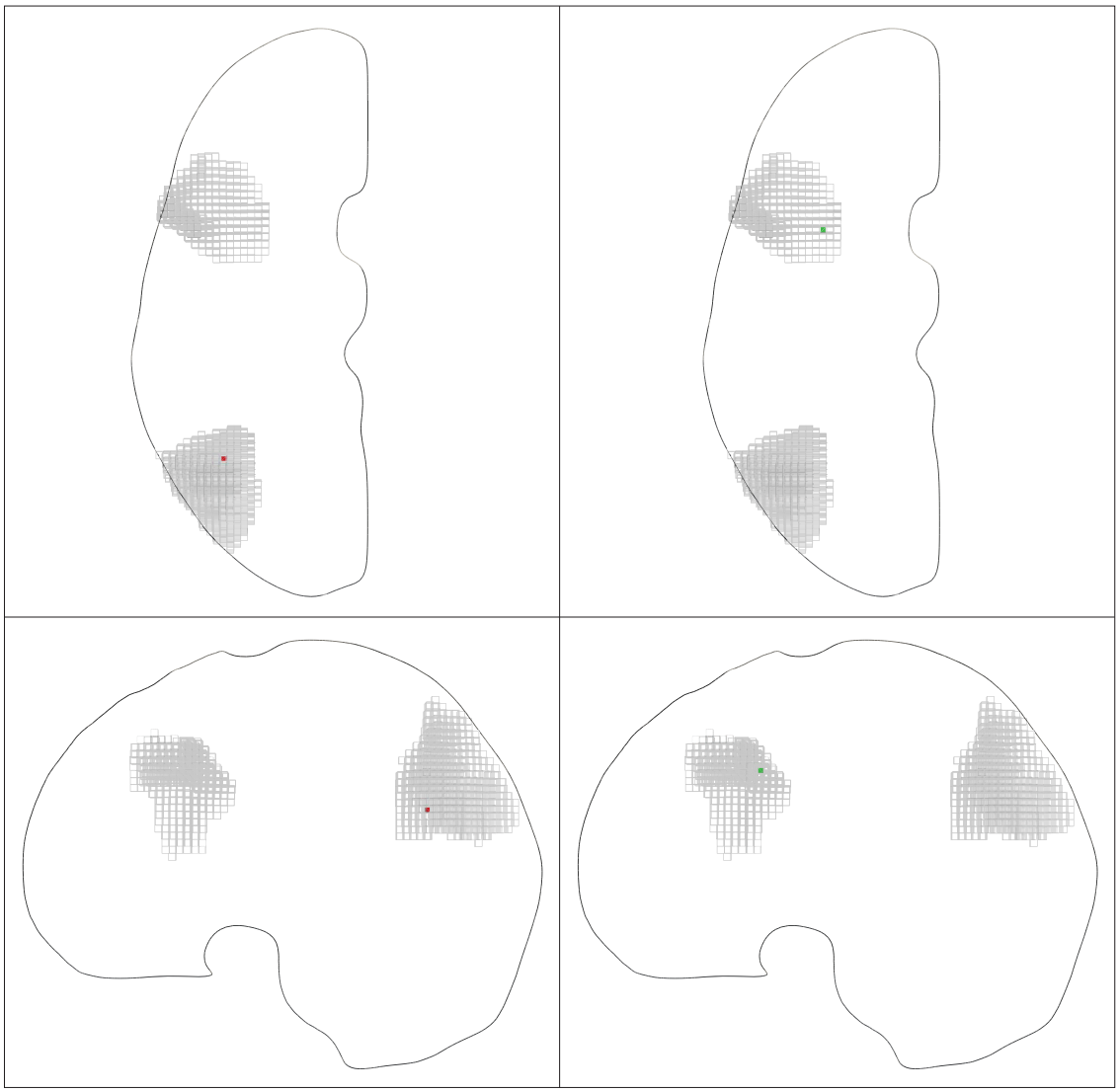


Figure 5.3 - Plots comparing univariate (left) and multivariate (right) models: peak.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that only one voxel was isolated. The Brodmann areas are represented as the grey wireframes.

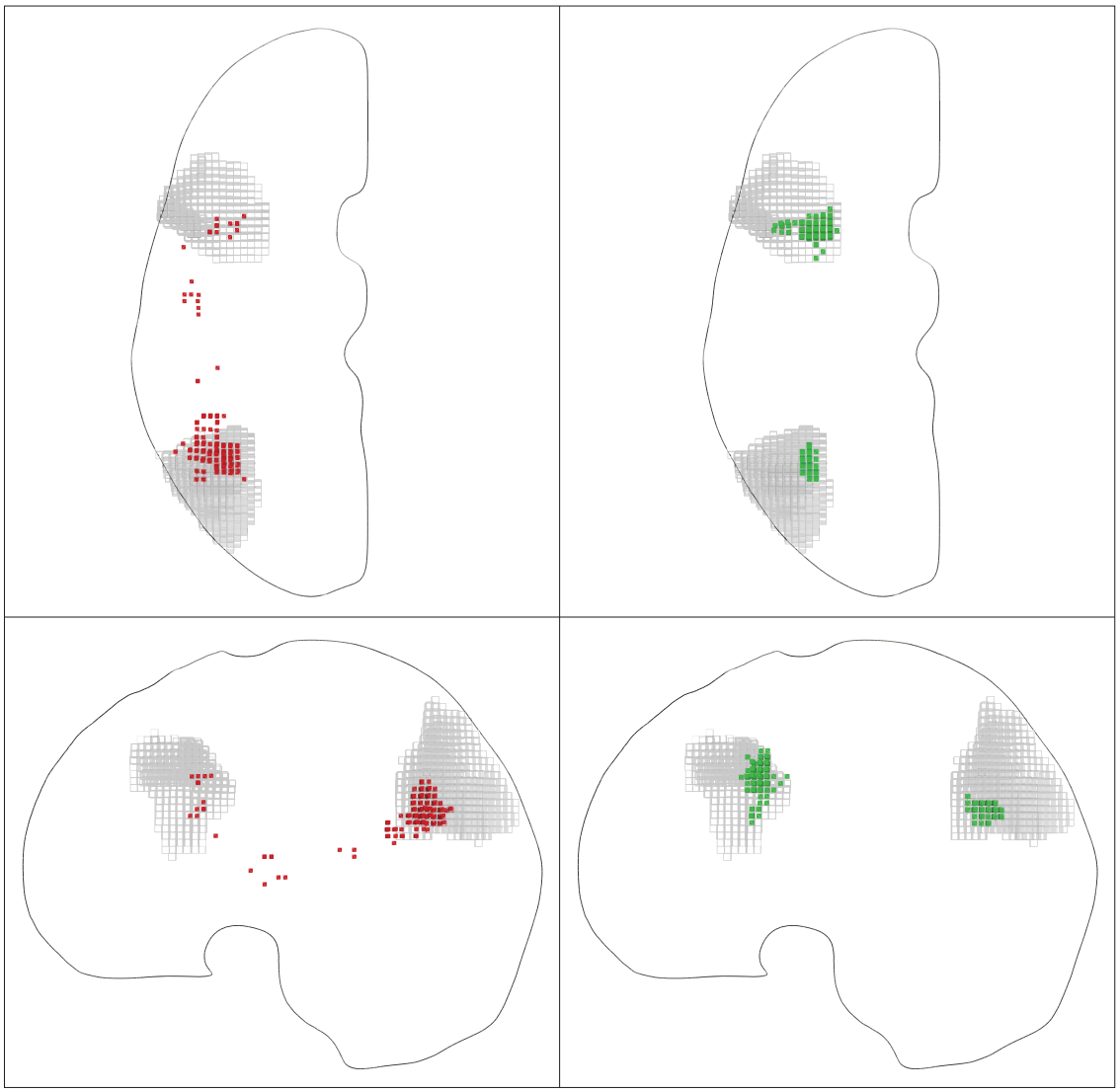


Figure 5.4 - Plots comparing univariate (left) and multivariate (right) models: 5%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 5% of the volume of BA39 or BA44 (grey wireframe).

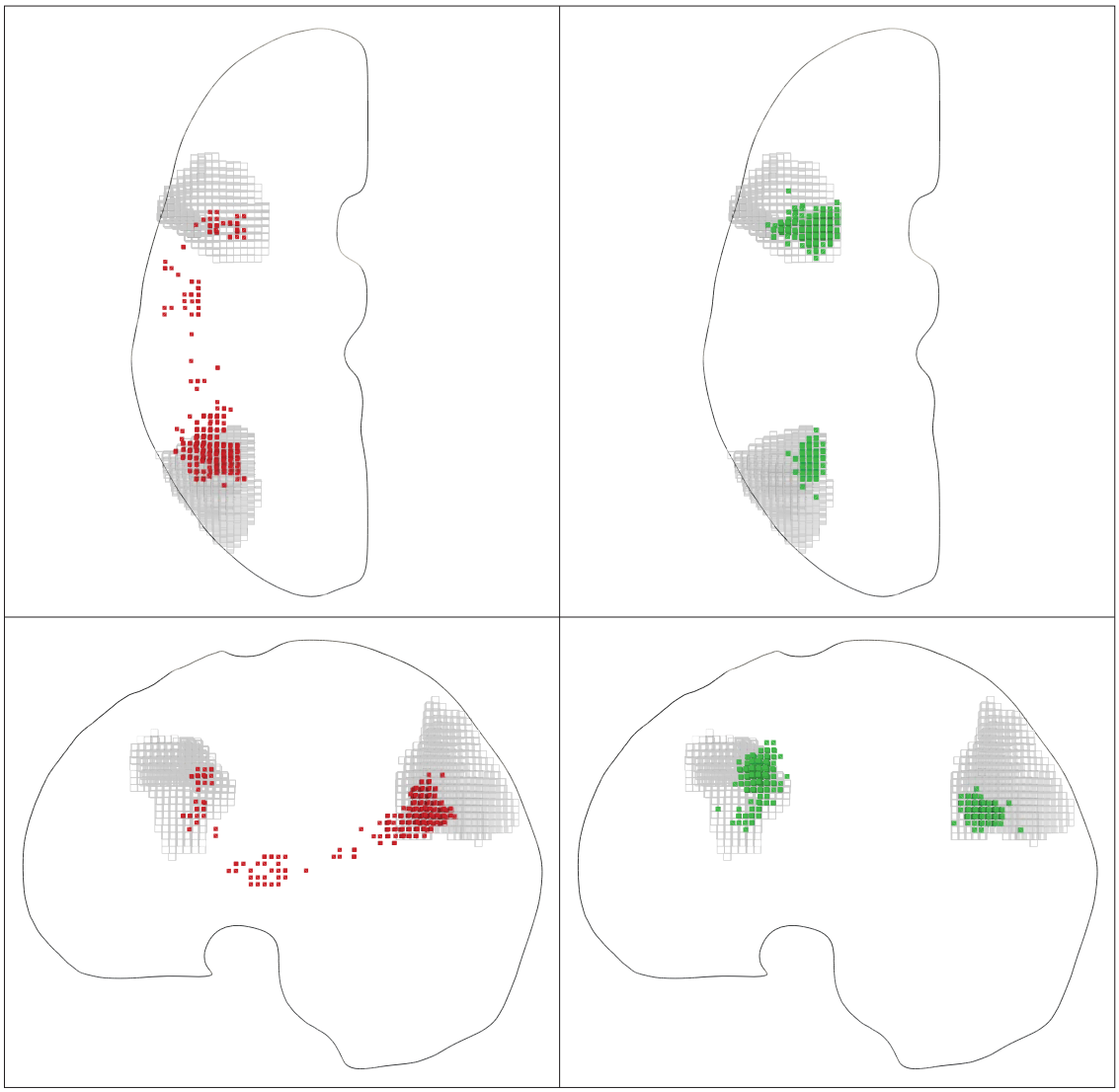


Figure 5.5 - Plots comparing univariate (left) and multivariate (right) models: 10%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 10% of the volume of BA39 or BA44 (grey wireframe).

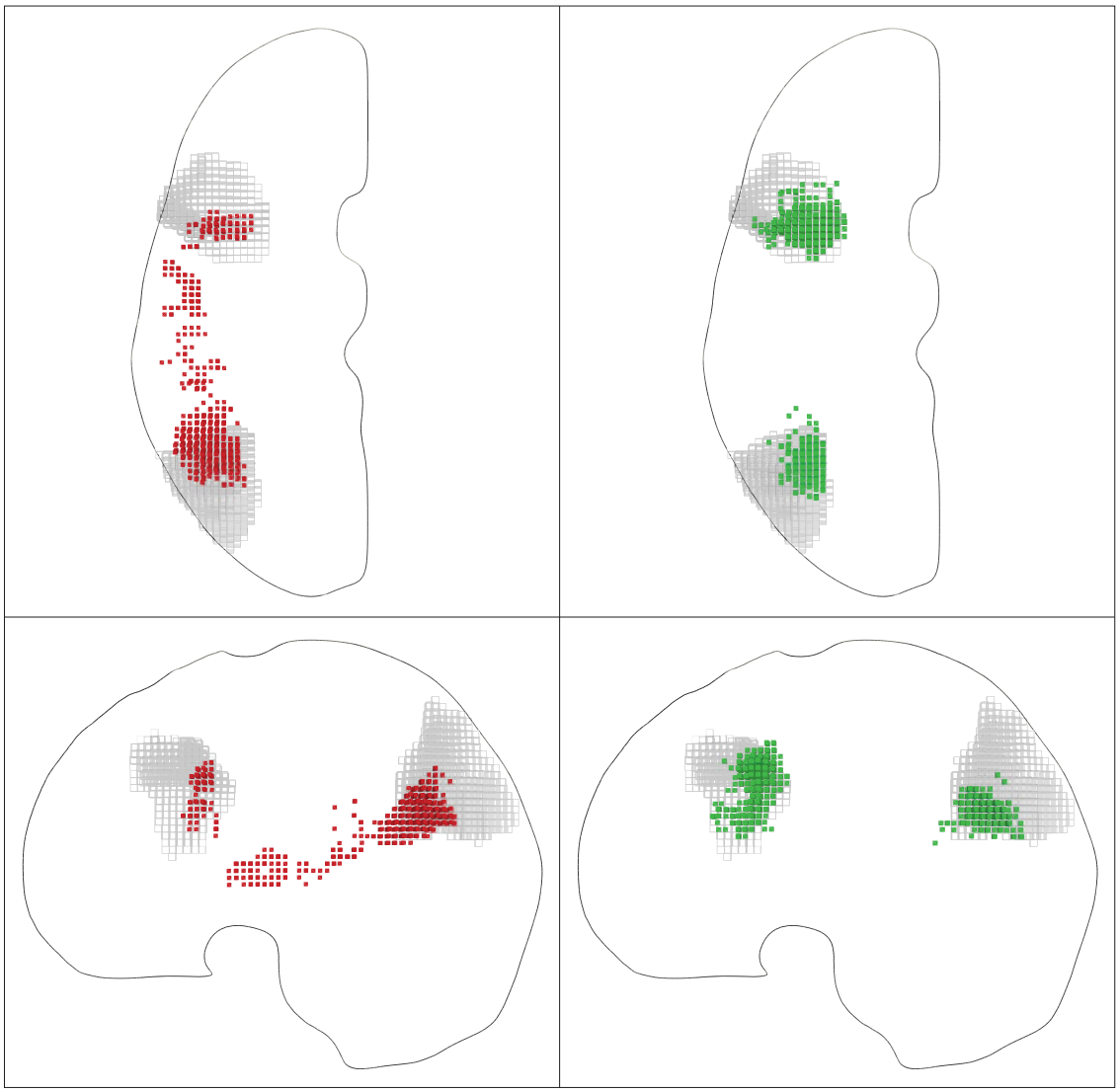


Figure 5.6 - Plots comparing univariate (left) and multivariate (right) models: 20%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 20% of the volume of BA39 or BA44 (grey wireframe).

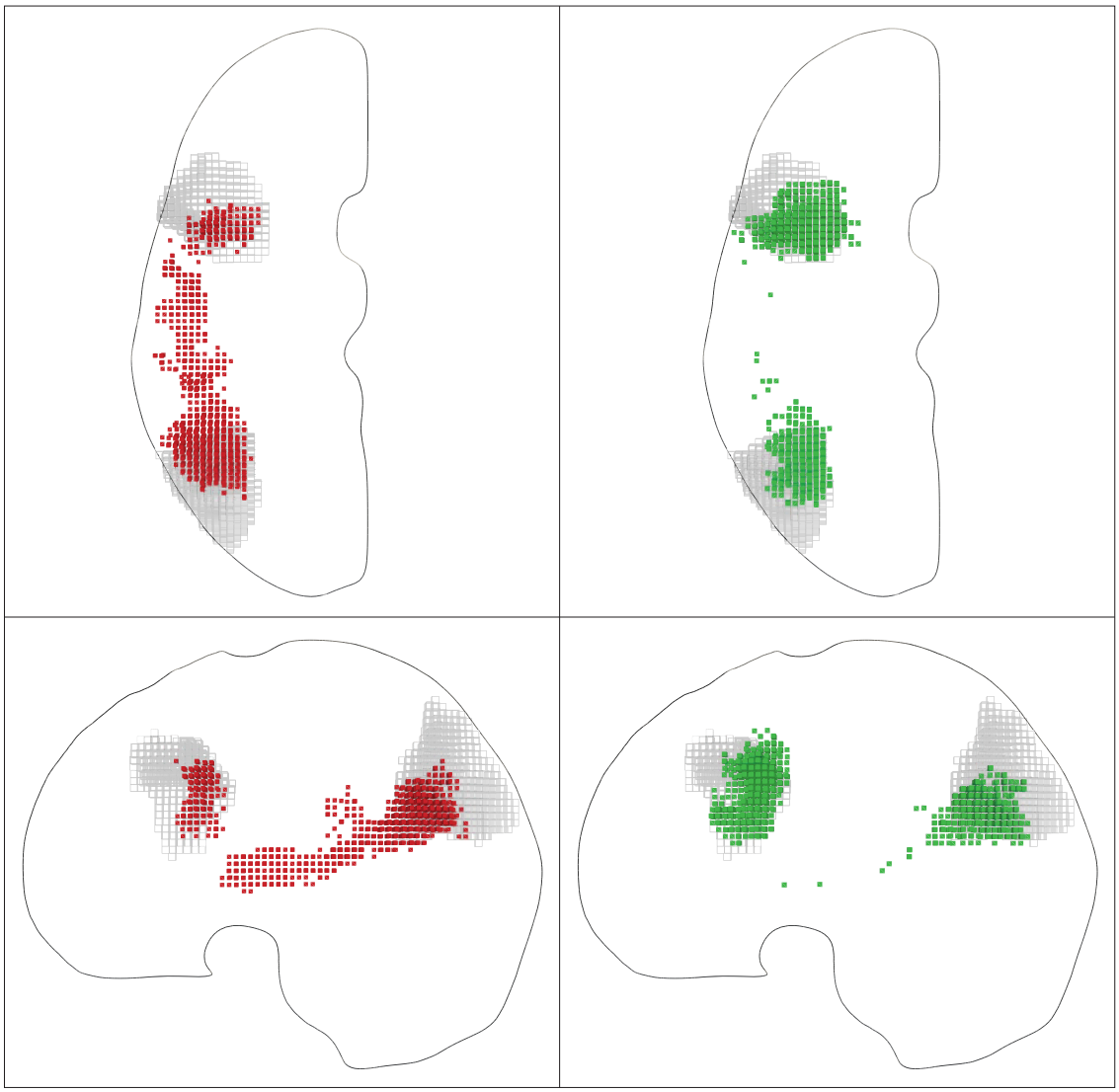


Figure 5.7 - Plots comparing univariate (left) and multivariate (right) models: 40%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 40% of the volume of BA39 or BA44 (grey wireframe).

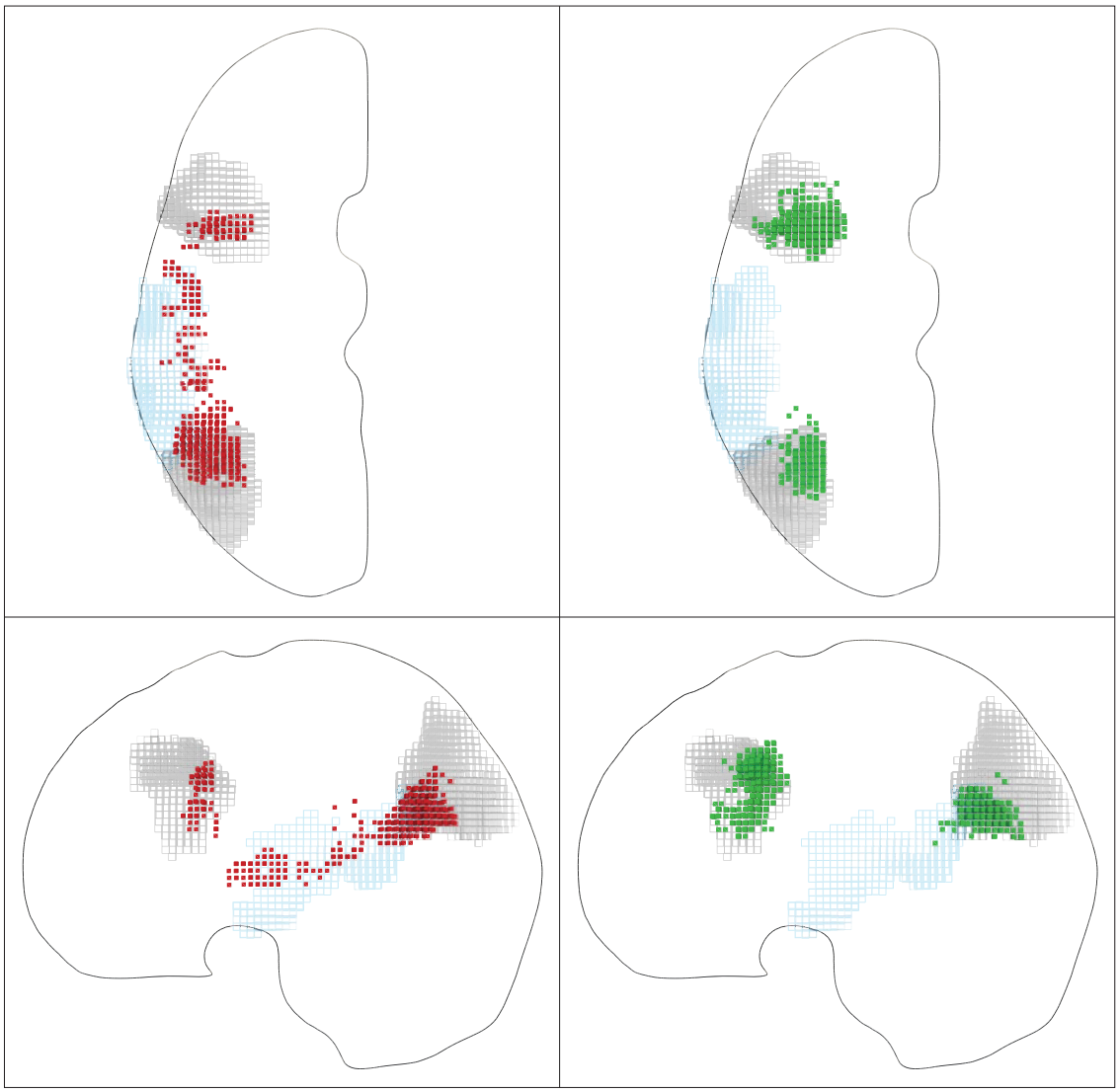


Figure 5.8 - Plots comparing univariate (left) and multivariate (right) models, with the superior temporal gyus outline: 20%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 20% of the volume of BA39 or BA44 (grey wireframe). Note that the centre of the significantly associated region for the mass univariate approach does not fall in either Brodmann area, but in the region of the superior temporal gyrus (blue wireframe). In comparison the mislocalisation is not seen in the SVM analysis.

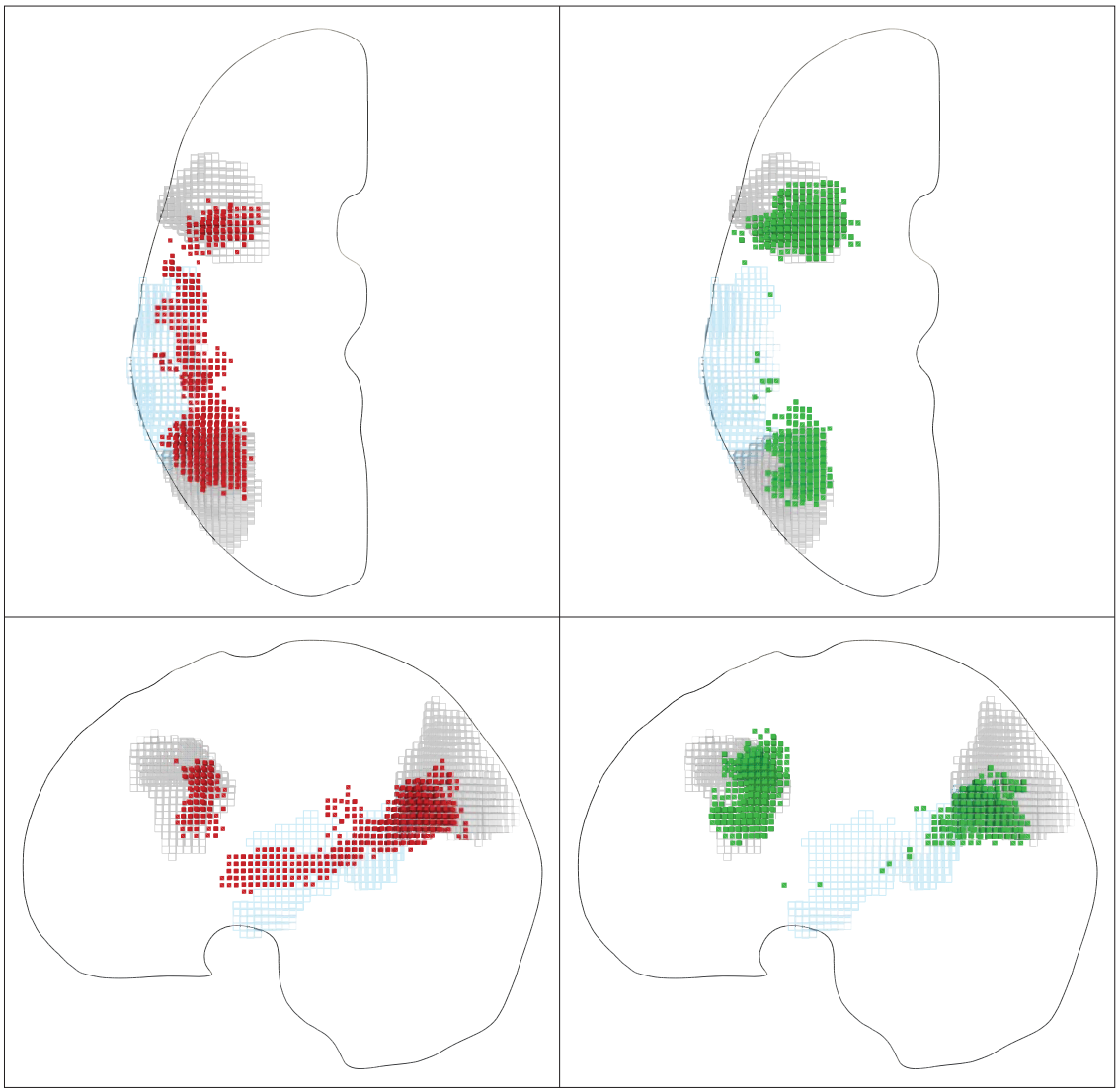


Figure 5.9 - Plots comparing univariate (left) and multivariate (right) models, with the superior temporal gyus outline: 40%.

Three dimensional plots of the areas identified as heavily weighted in the classification process, given damage to either Brodmann area 39 or Brodmann area 44 at greater than or equal to 20% of the volume, 90% of the time. A voxel-wise mass univariate analysis using the Fisher's exact test (red cubic glyphs – left column) and a Support Vector Machine analysis based on a linear kernel with a *C* value of 2^{-15} (green cubic glyphs – right column) of the sample of 581 acute stroke lesions. The *p* value brain map (mass univariate analysis) and the SVM weights brain map (SVM analysis) was thresholded at a level such that the volume of surviving voxels equalled 40% of the volume of BA39 or BA44 (grey wireframe). Note that the centre of the significantly associated region for the mass univariate approach does not fall in either Brodmann area, but in the region of the superior temporal gyrus (blue wireframe). In comparison the mislocalisation is not seen in the SVM analysis.

5.3.1.2 Multivariate analysis (SVM)

A search of the parameter space (*C*) was required prior to selecting the optimal model. A total of 41 different *C* values were assessed, with each processed with a 17 fold cross-validation of the data. The grid search assessment took 36 minutes, while the creation of the final model using the optimal *C* parameter value of 2^{-15} required only 5 seconds.

The optimal model based on a linear kernel with a *C* parameter of 2^{-15}, possessed a total of 394 support vectors (SVs) and a rho of -0.9975 indicating a bias within the data where "unaffected" cases are more likely. The range of weights was from -0.001572 to 0.0006047.

The dimension (voxel) with the greatest importance for the classification process is located at the posterior, superior aspect of BA44 (figure 5.3 – right column). At 5% (138 voxels), both BA39 and BA44 are populated by a cluster each along the posterior, superior border and anterior, inferior border respectively (figure 5.4 – right column).

In figures 5.3 through to 5.8 both these clusters continue to expand as the percentage volume of the true region is gradually increased from 10% through to 20%. There is some extension beyond the boundaries of the anterior, inferior border of BA39, however in contrast to the mass univariate analysis, there are no voxels populating the STG or space between the two critical regions (figures 5.8 and 5.9 – right column).

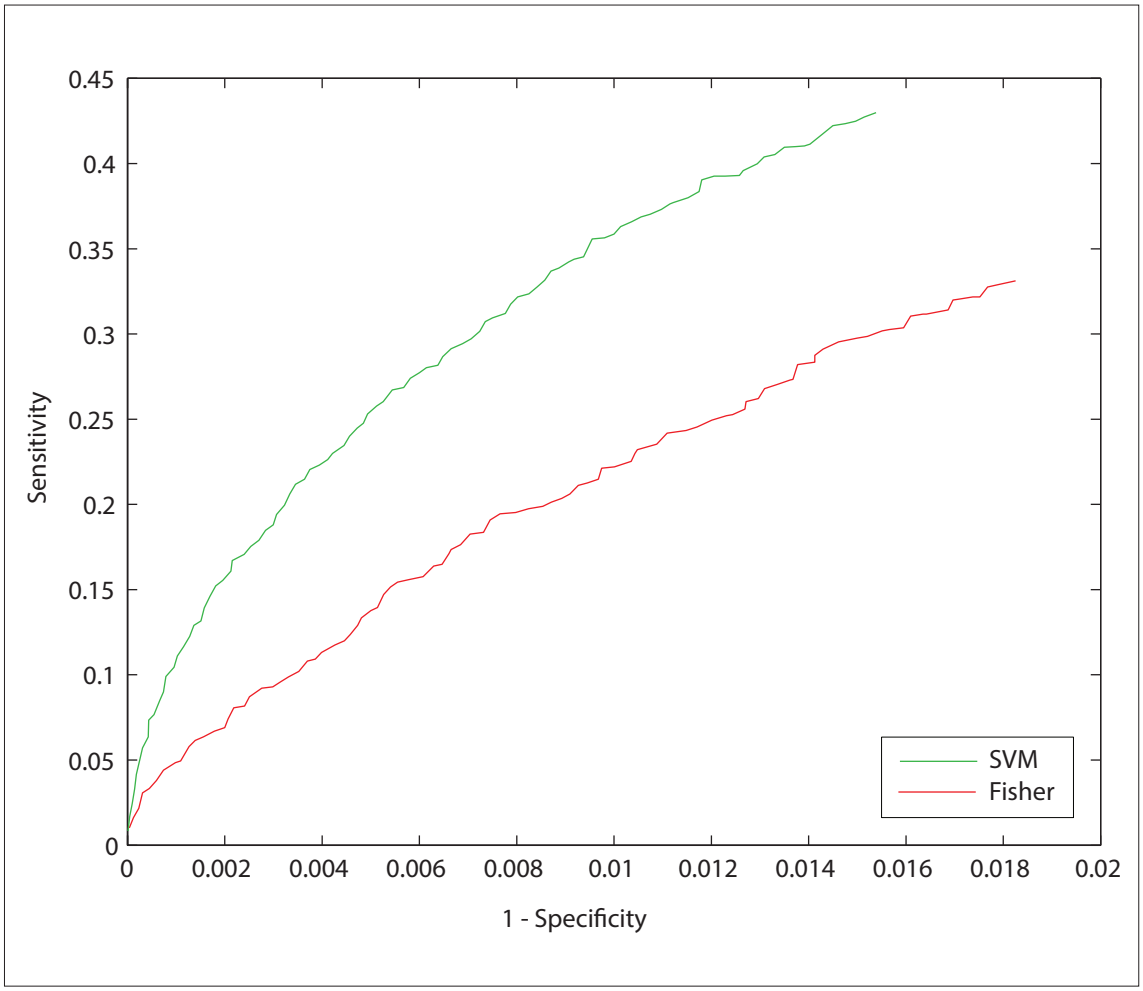


Figure 5.10 - Receiver operating curve comparing the univariate and multivariate models.

The sensitivity and specificity for the SVM and Fisher's exact mass univariate models is calculated for different fractions of the total "ground truth" volume specified by Brodmann area 39 and 44. The SVM model performs better across the entire volume range (1-100%) of the critical region.

5.3.2 Simulation two : Comparison of different multivariate techniques

The performance of each approach is assessed in three main parts. The first section provides details regarding the computation time of the method, which has implications on its future applicability. The next section examines the ability of the approach to identify the critical regions that constitute the hypothetical rule. This is achieved by extracting the derived associated voxel weights and rendering a 3 dimensional volume. Although the weights for the voxels have a polarity that determines the predictive characteristics for the voxel, the different approaches do not share the same magnitude scale. To facilitate the comparison between methods a common feature is needed, which in the following simulations is the rule determining the presence or absence of the dysfunction. The threshold necessary to select a specified fraction of the total critical volume – Brodmann areas 39 and 44 combined – can be identified for different fractions for each method. In this way comparisons between the fidelity of each method to the lesion symptom can be made. The third section examines the predictive performance of the generated models.

The hypothetical model used in this series of simulations was the dependence of a putative function to the injury of 30% or more of either Broadmann area 39 or 44 resulting in a hypothetical deficit of interest. If the lesion within the brain scan satisfied this criteria the scan would be classified as “affected”. 156 scans out of the total 581 images were kept separate from the training data. The predictive performance of each method was assessed on a balanced subset of 42 “affected” and “unaffected” scans from the 156 test set, as well as the entire test set.

5.3.2.1 Support Vector Machines (SVM)

The *C* parameter search explored the range between 2^{-20} and 2^{20}, with a 17 fold cross-validation (400 vs 25) assessment for each value. The duration of the parameter search across the 41 different *C* values took 2.5 minutes. The highest accuracy, using a linear kernel, was achieved with a *C* value of 2^{-9}.

The optimal model trained on 425 scans with a linear kernel and $C=2^{-9}$, possessed 92 support vectors (SVs), and a rho of -1.127. The range of weights was from -0.02626 to 0.01513.

Figure 5.11 and figure 5.15 (top row) shows the dimension (voxel) with the greatest importance for the classification process is located at superior aspect of BA44. By 5% of the total true volume (17 voxels), BA39 has not been populated, with only the cluster in BA44 continuing to expand.

At 10% of the true volume (figures 5.12 and 5.15 – top row), a cluster appears in the centre of BA39. A single voxel appears between the 2 critical regions at 15% of the true volume (52 voxels).

In figures 5.11 through to 5.16, as the percentage of the true volume is gradually increased up to 50% of the true volume; both the clusters within BA39 and BA44 continue to expand. There is some extension outside the border of BA44 along the superior, inferior axis, and the inferior border of BA39. The single voxel between BA39 and BA44 does become more populated, but, the cluster remains very sparse, and does not merge with either BA39 and BA44.

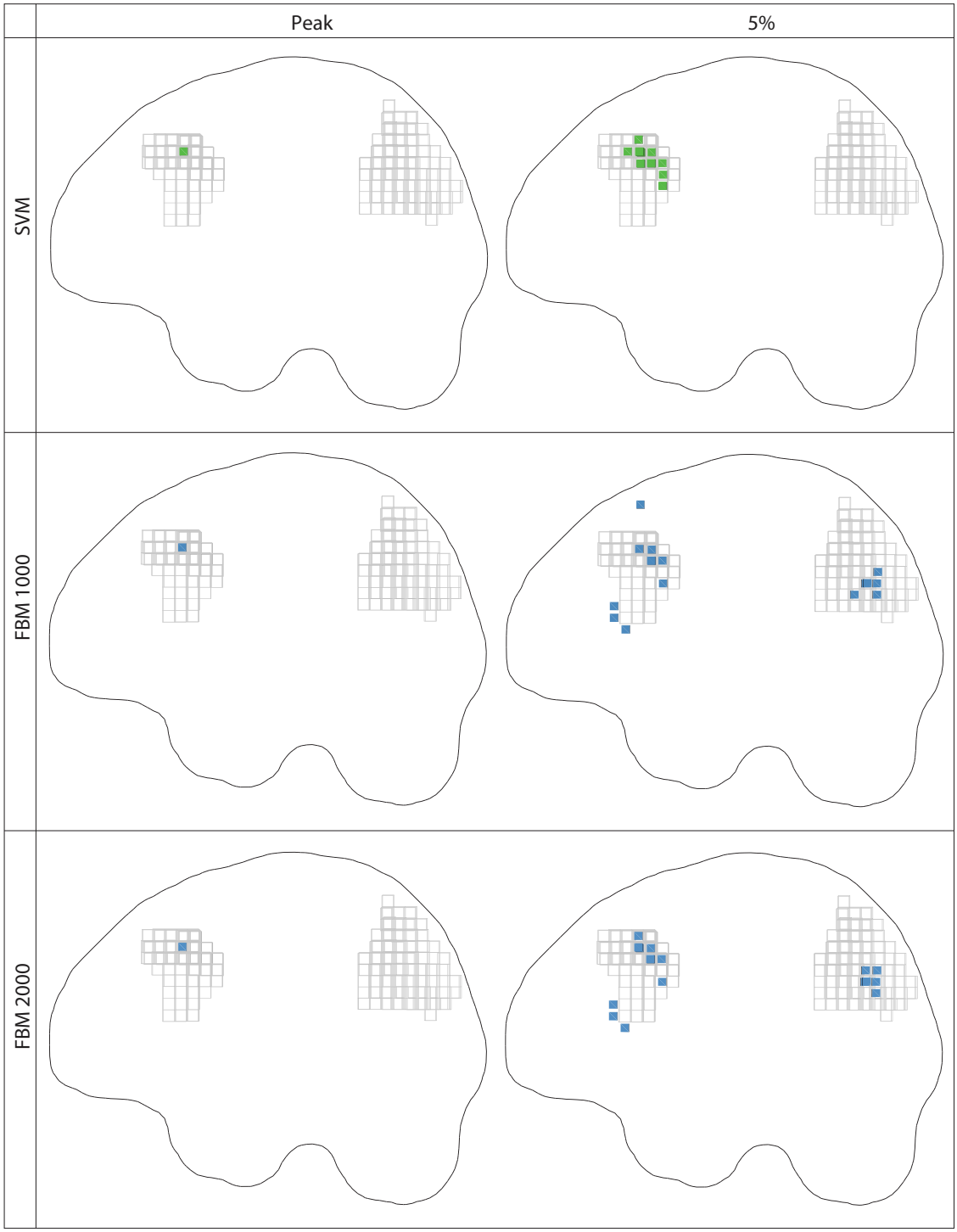


Figure 5.11 - Plots comparing SVM and FBM models: peak and 5%.

Three dimensional plots, in sagittal section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled the peak and 5% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs) and a full Bayesian MCMC model with either 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

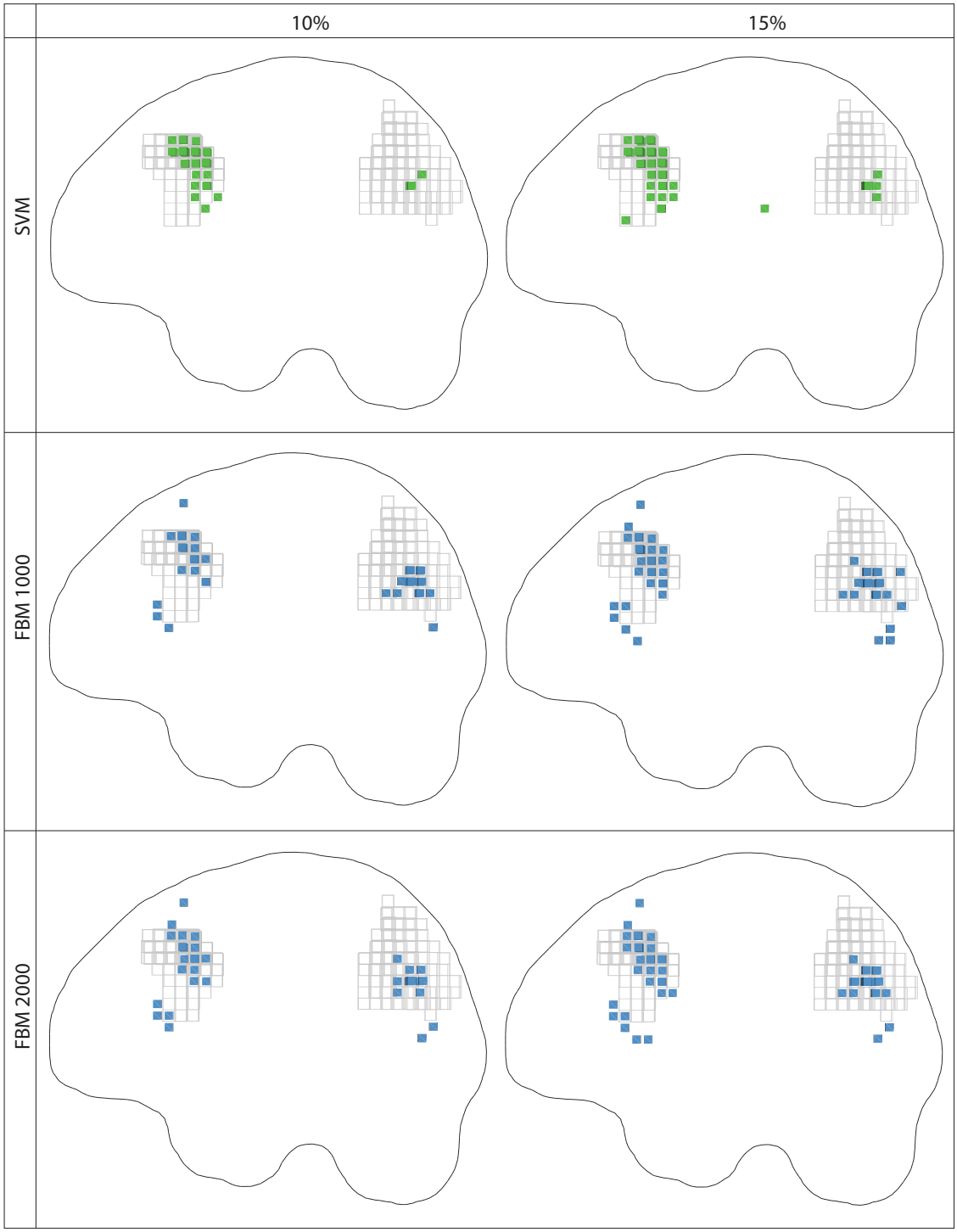


Figure 5.12 - Plots comparing SVM and FBM models: 10% and 15%.

Three dimensional plots, in sagittal section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled 10% and 15% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs) and a full Bayesian MCMC model with either 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

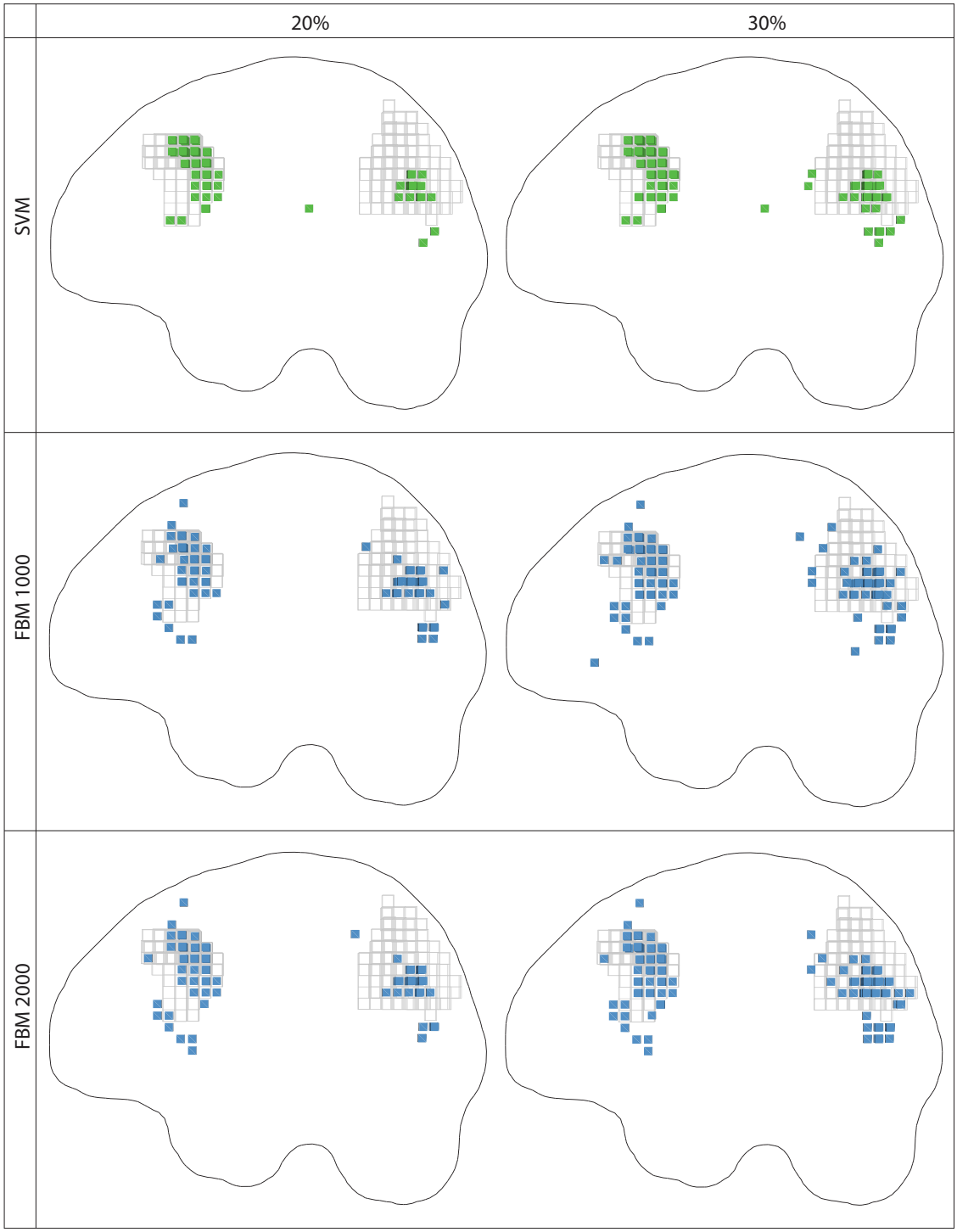


Figure 5.13 - Plots comparing SVM and FBM models: 20% and 30%.

Three dimensional plots, in sagittal section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled 20% and 30% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs) and a full Bayesian MCMC model with either 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

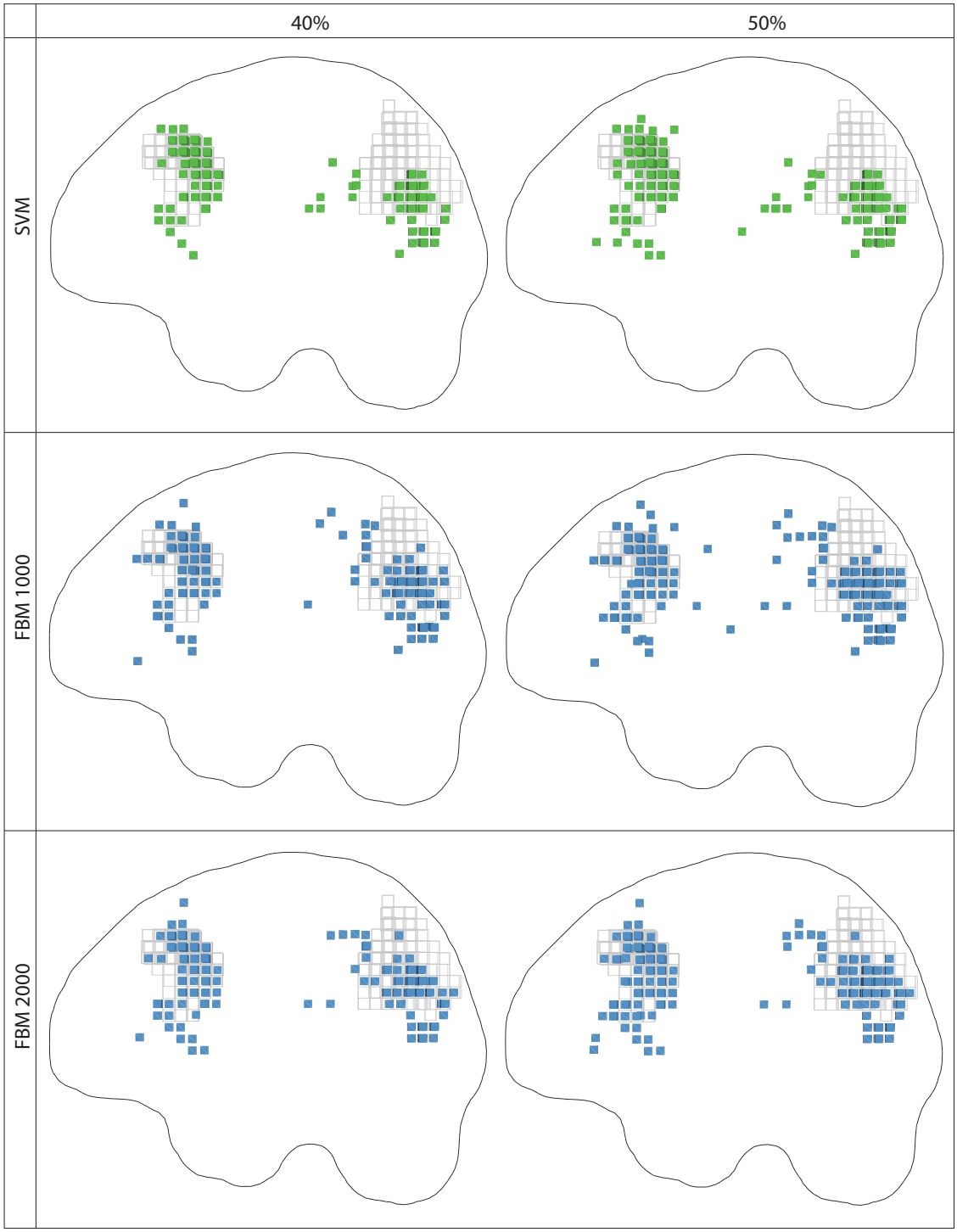


Figure 5.14 - Plots comparing SVM and FBM models: 40% and 50%.

Three dimensional plots, in sagittal section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled 40% and 50% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs) and a full Bayesian MCMC model with either 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

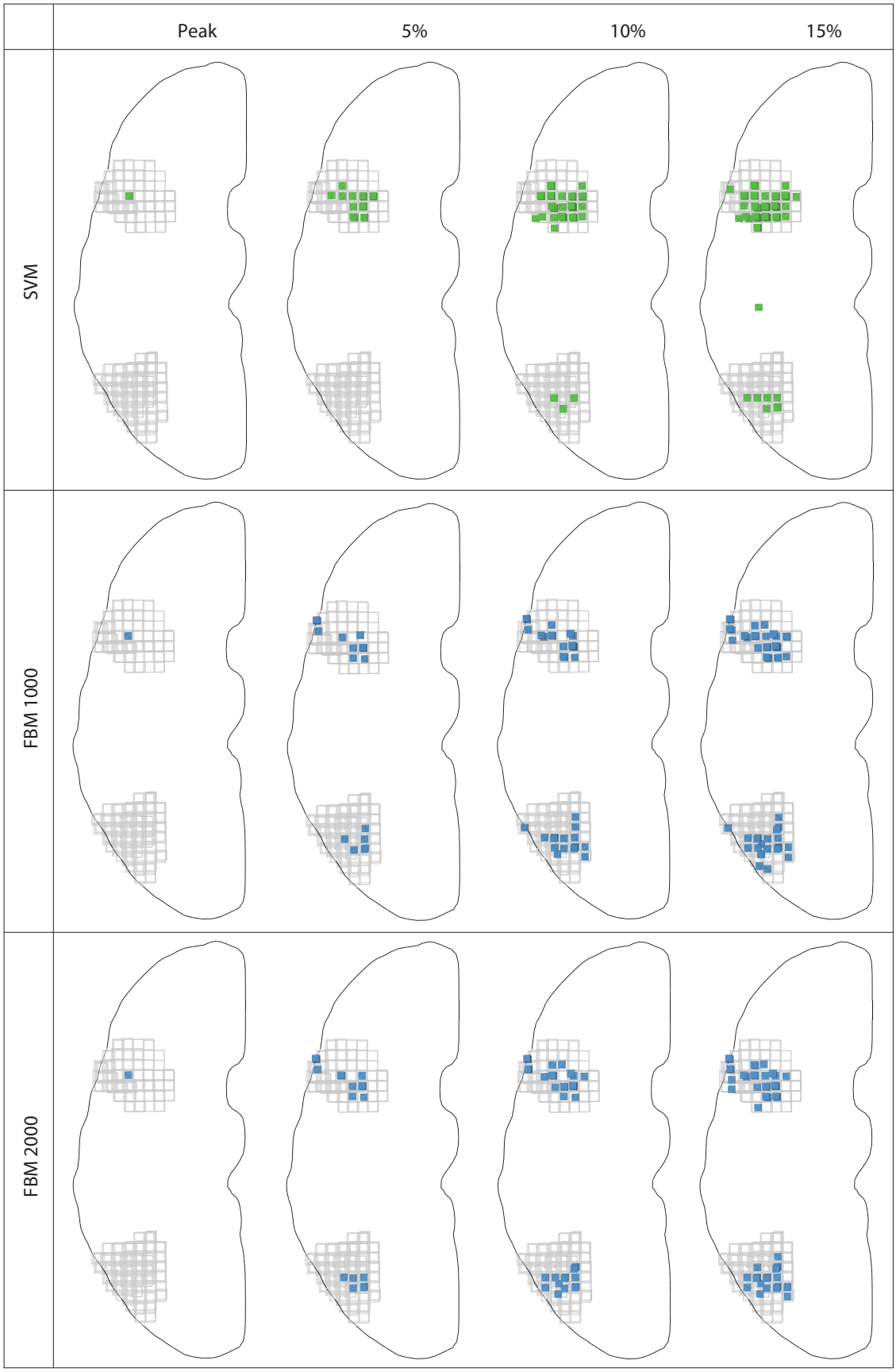


Figure 5.15

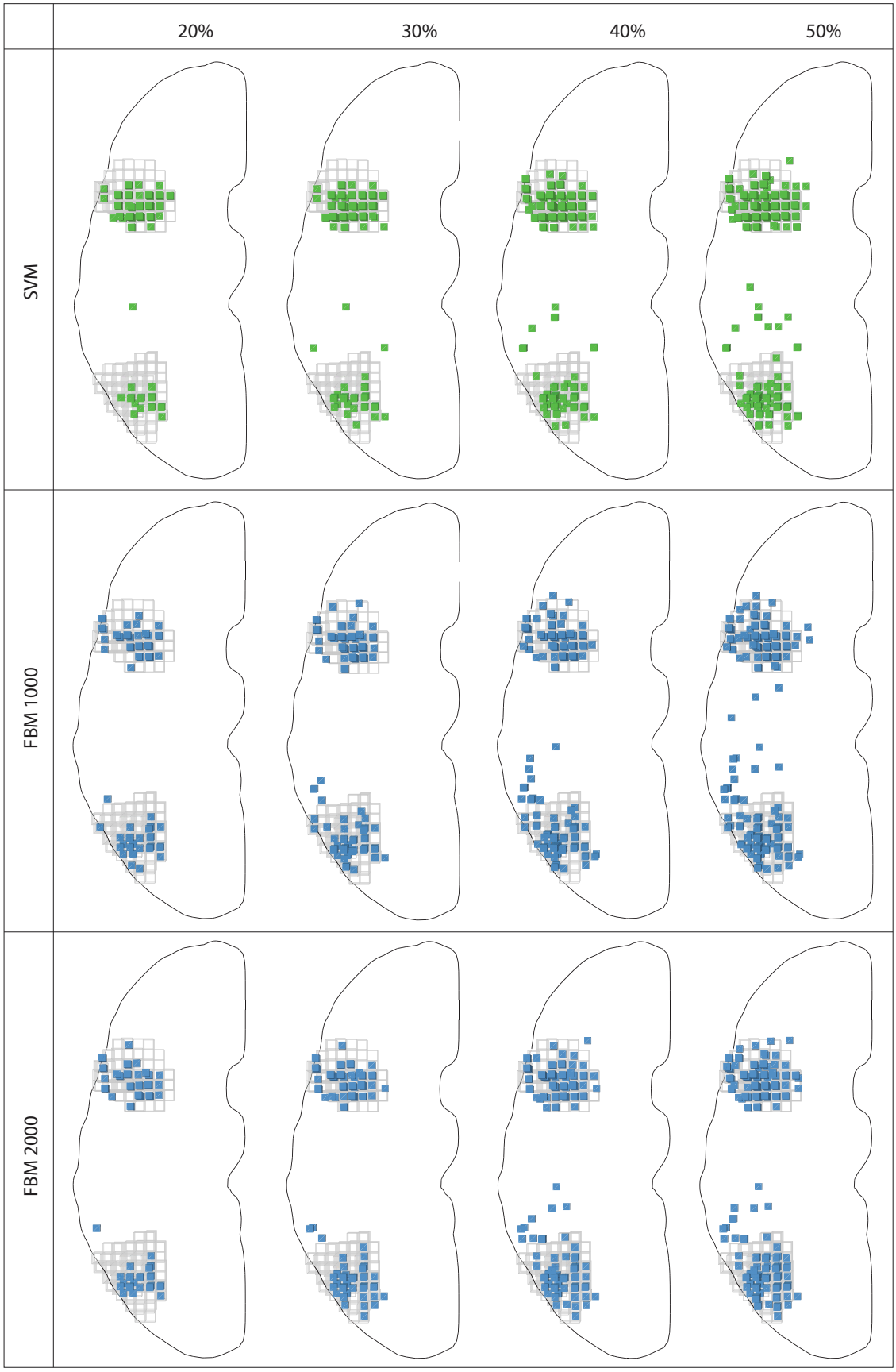


Figure 5.16

Figure 5.15 - Plots comparing SVM and FBM models: peak, 5%, 10%, 15%.

Three dimensional plots, in axial section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled the peak, 5%, 10% and 15% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs) and a full Bayesian MCMC model with 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

Figure 5.16 - Plots comparing SVM and FBM models: 20%, 30%, 40%, 50%.

Three dimensional plots, in axial section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The brain map of weights was thresholded such that the volume of surviving voxels equalled 20%, 30%, 40% and 50% of the volume of BA39 and BA44 (grey wireframe). The three different models are a SVM model based on a linear kernel (green cubic glyphs); a full Bayesian MCMC model with either 1000 or 2000 iterations and a burn in period of 500 or 1000 respectively (blue cubic glyphs).

5.3.2.2 Relevance Vector Machines (RVM)

A total of 212 different RVM models were created from the training set of 425 brains. The mean length of time to create one of these models was 92.74 seconds, with a range of 34.73 to 416.3 seconds. The number of relevance vectors identified by the algorithm ranged from 5 to 62, with a mean number of 57.19. The average minimum and maximum weights for the 212 different models were -5.001 and 15.38 respectively.

To compensate for the sparseness of the model, the mean weight for each dimension was calculated from the 212 model iterations and then mapped into their corresponding locations within the brain volume. A total of 367 unique RVs (74 positive and 293 negative) were identified.

The performance of the RVM model at identifying BA39 and BA44 was poor. The dimension (voxel) with the largest mean positive value is not located within either of the 2 Brodmann areas. It is positioned in the occipital lobe, posterior and medial to BA39 (figures 5.17 and 5.18 – top right panel). At 5% of the total number of positive RVs there is only one voxel which is on the external anterior, inferior border of BA44, whilst the other 3 RVs are located outside the true volume. Neither BA39 or BA44 are populated by any of the 74 positive (nor any of the negative) RVs, with a preponderance to populate the inferior half of the brain (figures 5.17 and 5.18 – bottom left panel).

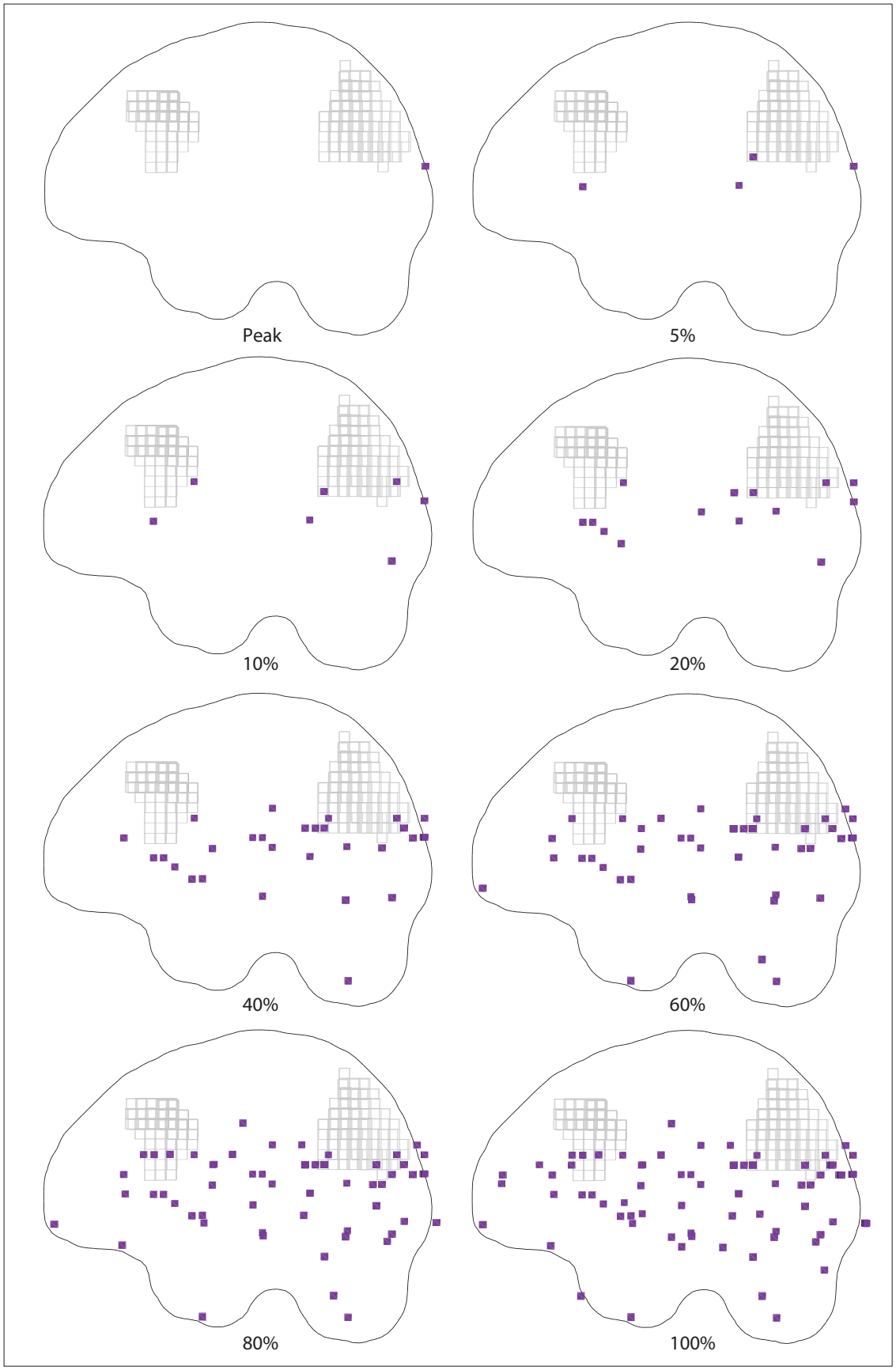


Figure 5.17

Figure 5.17 - Plot displaying the RVM model: sagittal plane.

Three dimensional plots, in sagittal section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The Relevance Vector Machine was iterated 212 on the sample of 425 scans (423 vs 2). The relevance vectors (RVs) were summed across the 212 models and the mean weights for the 74 unique positive RVs calculated. The RVs were then thresholded such that the volume of surviving voxels equalled the peak, 5%, 10%, 20%, 40%, 60%, 80% and 100% of the total number of unique positive RVs.

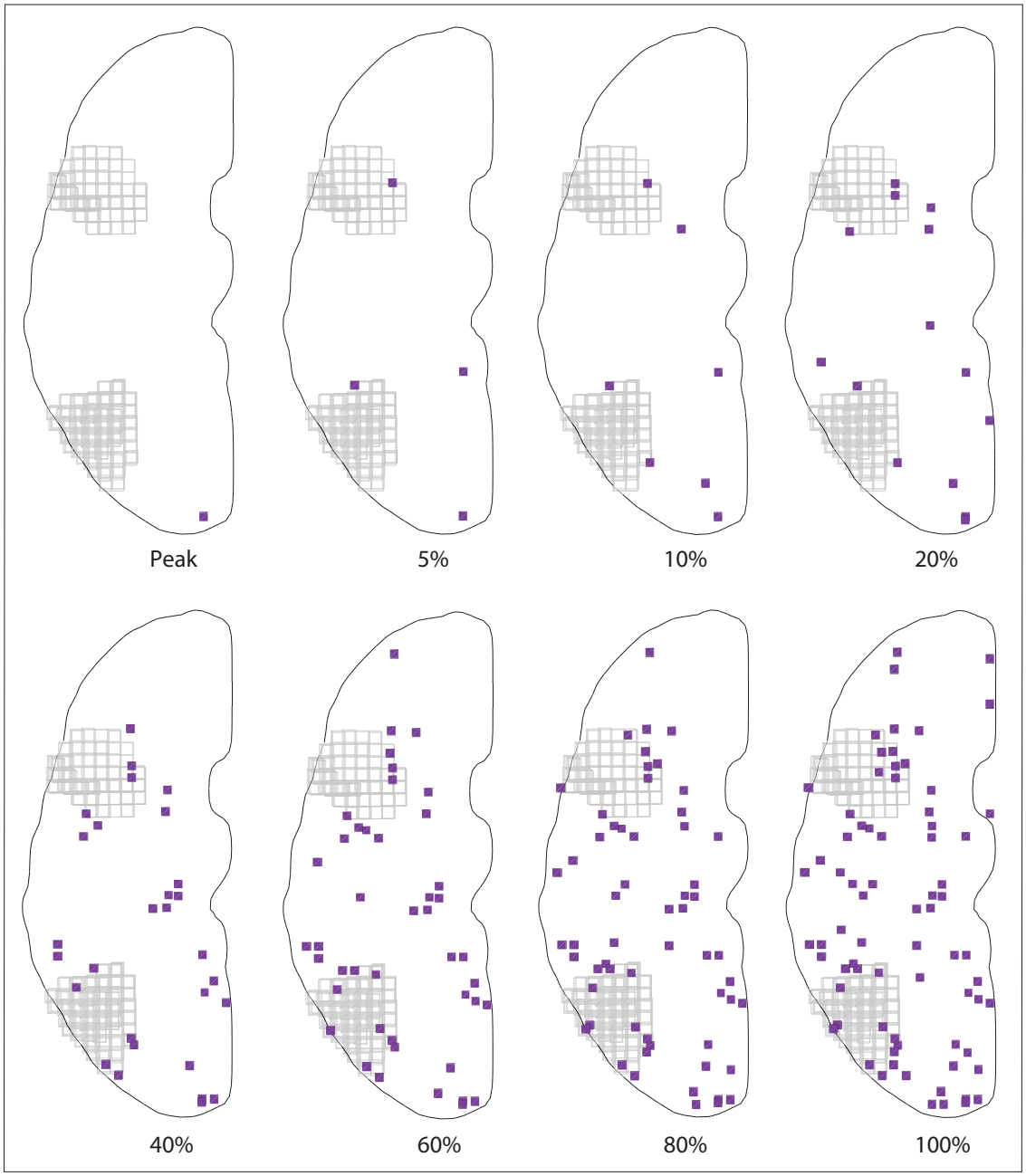


Figure 5.18 - Plot displaying the RVM model: axial plane.

Three dimensional plots, in axial section, of the areas identified as heavily weighted in the classification process, given damage to either BA39 or BA44 at greater than or equal to 30% of their total volume, by a high dimensional multivariate analysis of the sample 425 acute stroke lesions. The Relevance Vector Machine was iterated 212 on the sample of 425 scans (423 vs 2). The relevance vectors (RVs) were summed across the 212 models and the mean weights for the 74 unique positive RVs calculated. The RVs were then thresholded such that the volume of surviving voxels equalled the peak, 5%, 10%, 20%, 40%, 60%, 80% and 100% of the total number of unique positive RVs.

5.3.2.3 Flexible Bayesian Modelling

This multivariate method was the most computationally demanding, requiring 3-6 weeks to generate a predictive model, depending on the number of iterations specified. Since it is unclear what the optimal number of iterations and length of burn in period is supposed to be, 2 simulations were run where the model was created from 1000 iterations and a burn in period of 500, and another with 2000 iterations and a burn in period of 1000.

5.3.2.3.1 Burn in period

The weight associated with each dimension (voxel) fluctuates from iteration to iteration. In the ideal scenario the iteration to iteration fluctuation will gradually decrease as the Markov chain approaches stationarity. Figures 5.19 and 5.20 display the autocorrelation function plot for 12 randomly selected voxels over 1000 and 2000 iterations respectively. The selected voxels show stationarity being approached between 400 and 500 iterations in both cases, thus tending towards a random sampling behaviour by 400-500 iterations and achieving a good estimate of the posterior probability.

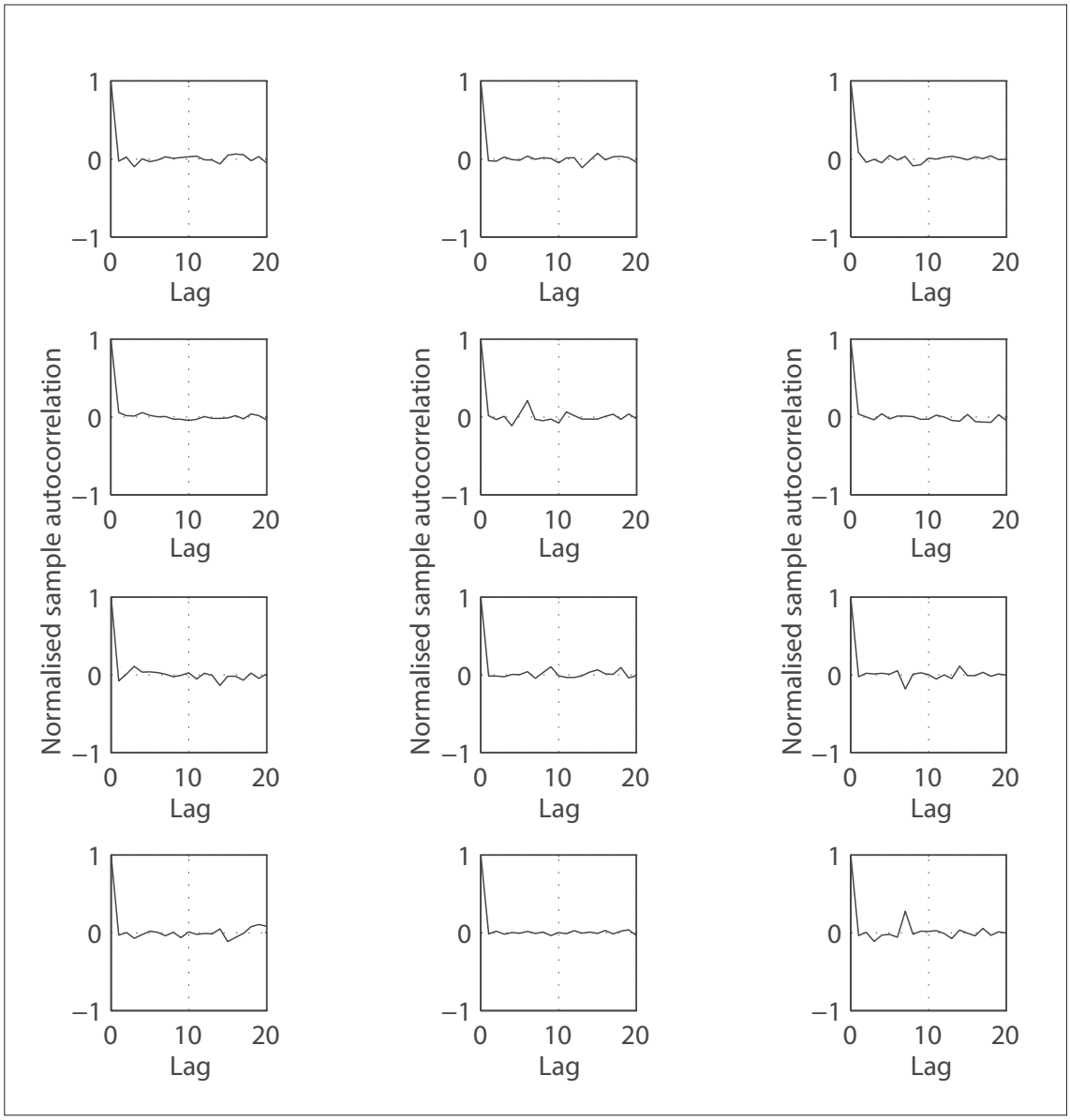


Figure 5.19 - Autocorrelation functions for 12 randomly selected voxels from the FBM 1000 iteration model.

The weight of every dimension in this FBM model is estimated iteratively 1000 times. The following panels show the autocorrelation functions for 12 voxels. Each lag interval contains 50 iterations and the autocorrelation function has been normalised such that at lag zero the autocorrelation function (ACF) is equal to 1. In all 12 voxels the amount of cross-correlation drops significantly after 400 iterations.

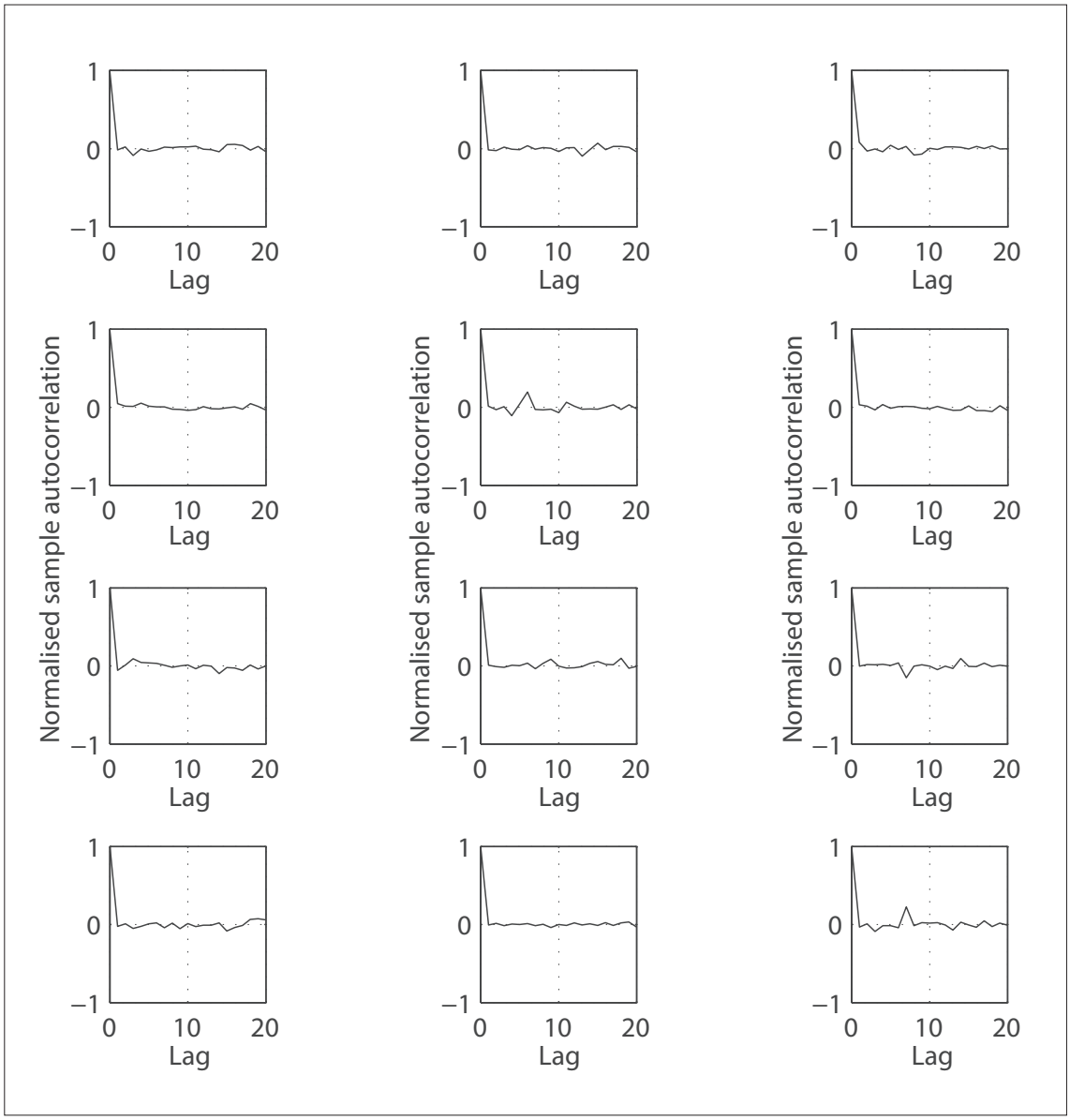


Figure 5.20 - Autocorrelation functions for 12 randomly selected voxels from the FBM 2000 iteration model.

The weight of every dimension in this FBM model is estimated iteratively 2000 times. The following panels show the autocorrelation functions for 12 voxels. Each lag interval contains 100 iterations and the autocorrelation function has been normalised such that at lag zero the autocorrelation function (ACF) is equal to 1. In all 12 voxels the amount of cross-correlation drops significantly after 400 iterations.

Although each dimension will have a different weight, the mean weight for the entire brain volume should tend towards a fixed value. Figures 5.21 and 5.22 show the trend for the mean weight of the entire hemisphere against iteration number, while figure 5.23 and 5.24 show the autocorrelation function for the entire hemisphere. Both plots suggest that stationarity is being approached by 500. As long as the process appears to be heading towards stationarity, theoretically the accuracy of the weight ascribed by the model should improve with increasing iterations. Pragmatically the use of a burn in period of 500 and 1000 respectively for the model with 1000 iterations and 2000 iterations were used for the visualisations and predictions.

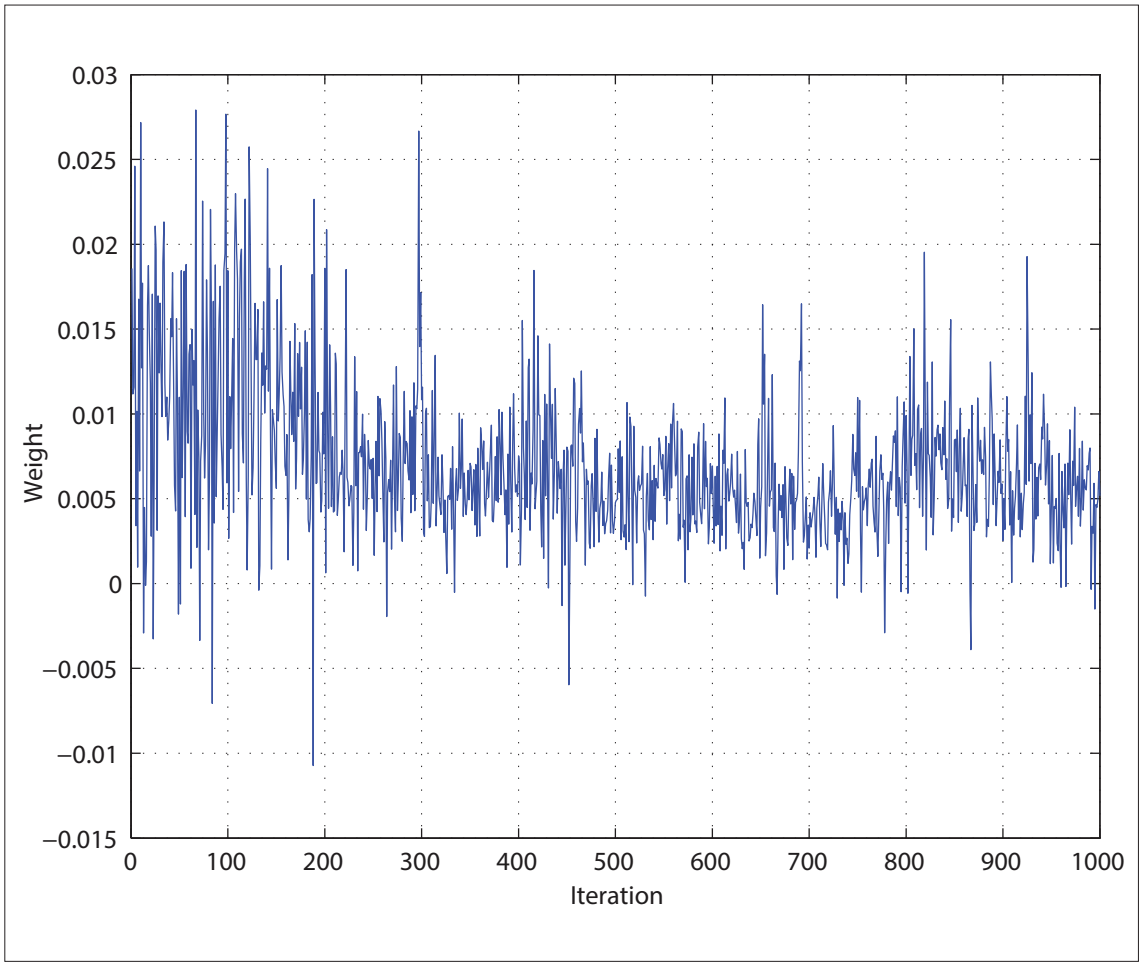


Figure 5.21 - Mean trace plot for the FBM 1000 iteration model.

The weight of every dimension in the FBM model is estimated iteratively. The following plot shows the mean trace for all voxels from the FBM model and 1000 iterations.

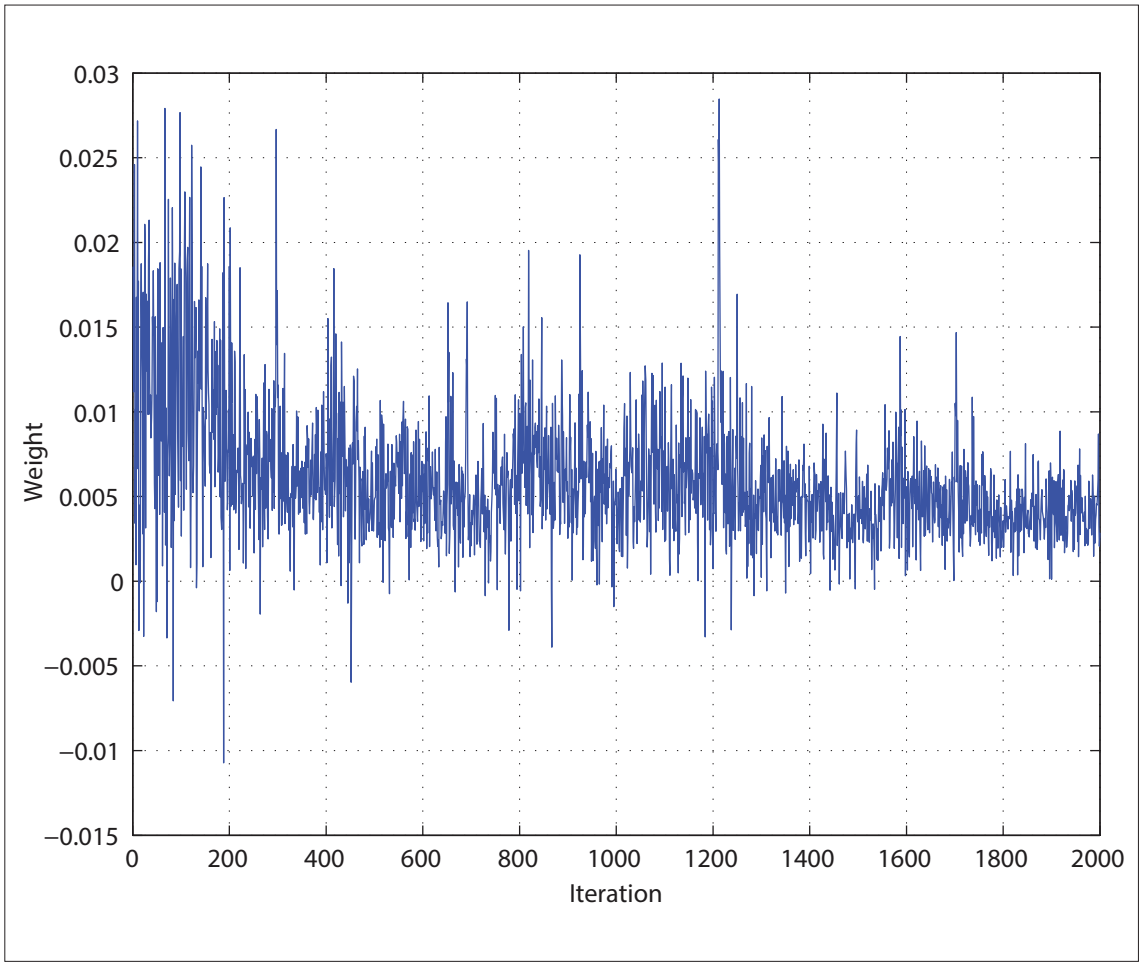


Figure 5.22 - Mean trace plot for the FBM 2000 iteration model.

The weight of every dimension in the FBM model is estimated iteratively. The following plot shows the mean trace for all voxels from the FBM model and 2000 iterations.

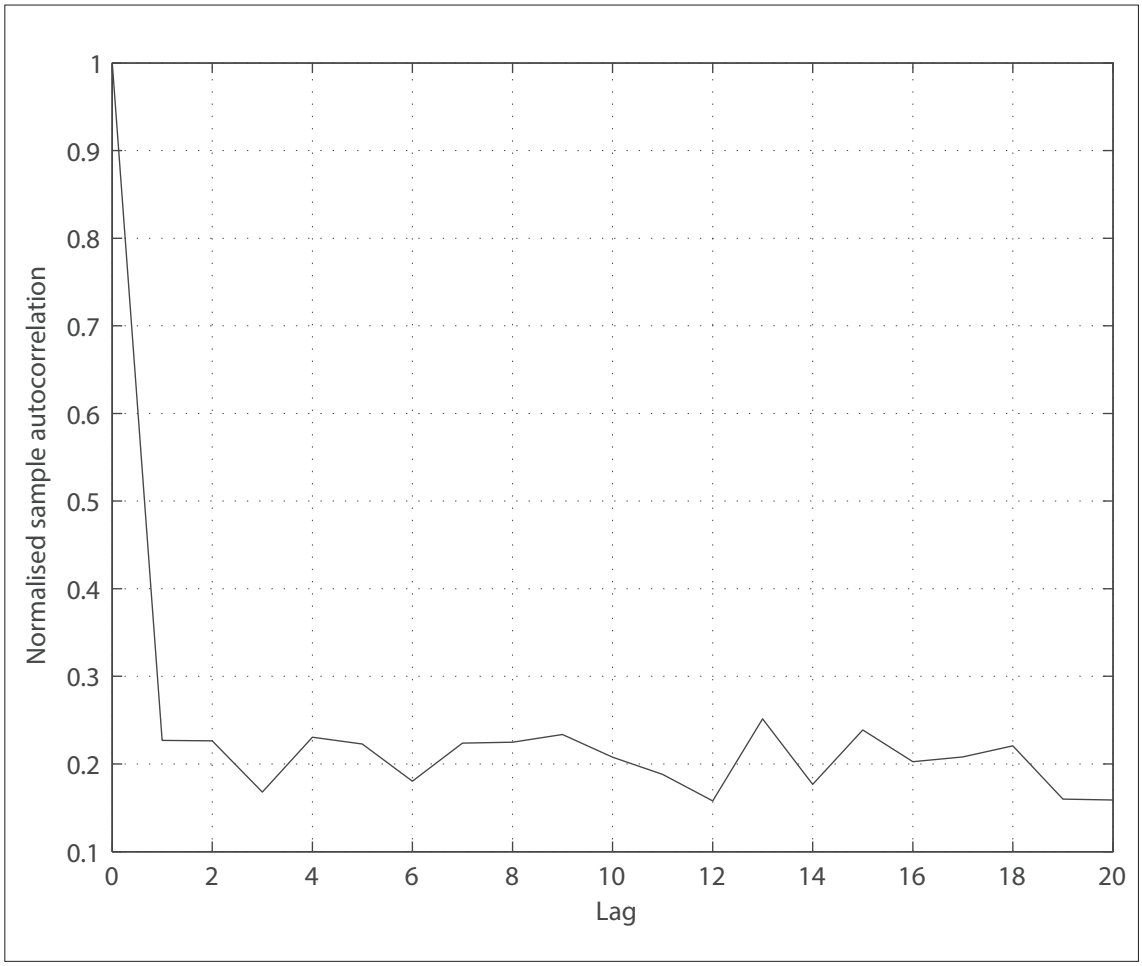


Figure 5.23 - Autocorrelation function for the FBM 1000 model.

The weight of every dimension in this FBM model is estimated iteratively 1000 times. The following plot shows the mean normalised autocorrelation function (ACF) for all voxels. Each lag interval contains 50 iterations and the autocorrelation function has been normalised such that at lag zero the ACF is equal to 1.

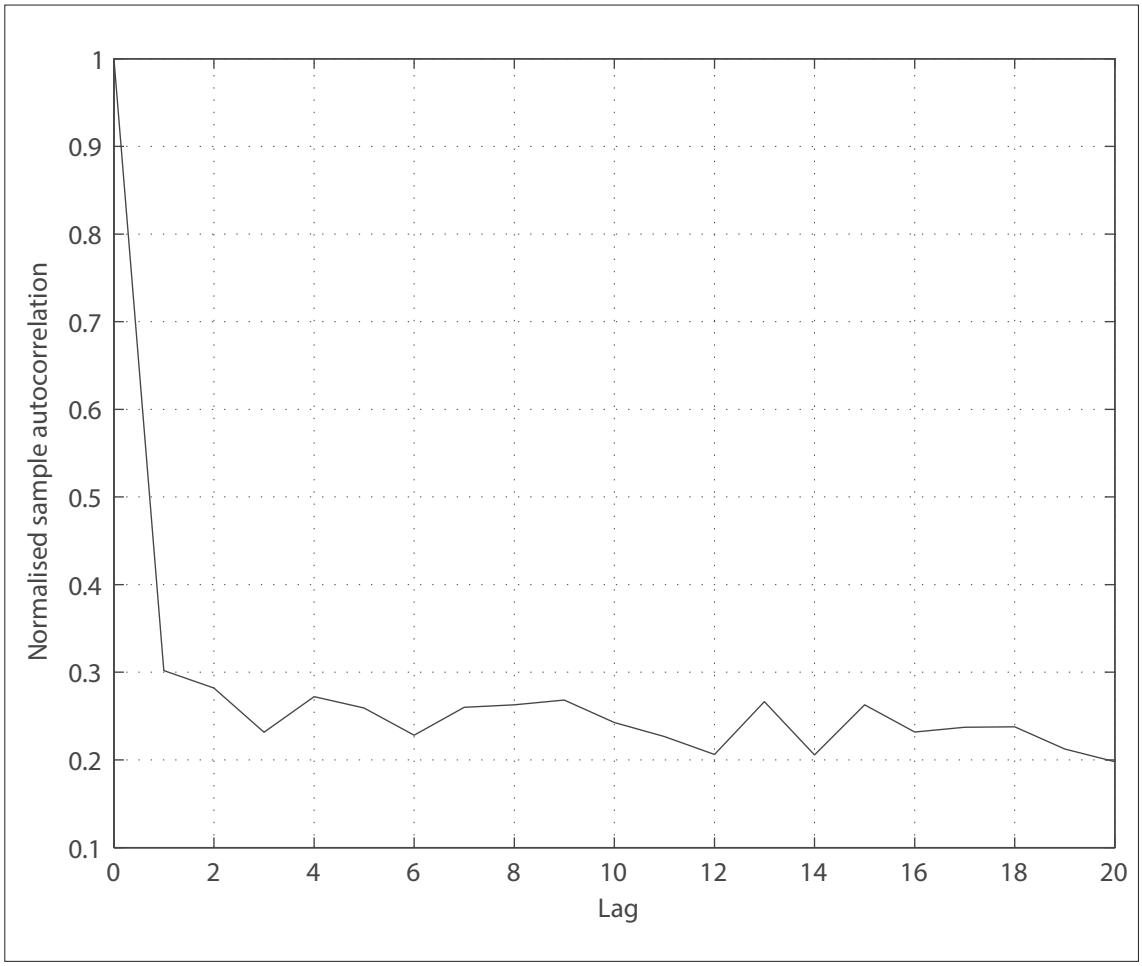


Figure 5.24 - Autocorrelation function for the FBM 2000 model.

The weight of every dimension in this FBM model is estimated iteratively 2000 times. The following plot shows the mean normalised autocorrelation function (ACF) for all voxels. Each lag interval contains 100 iterations and the autocorrelation function has been normalised such that at lag zero the ACF is equal to 1.

The range of weights for the 1000 model and 2000 model are very similar at -0.1009–3.387 and -0.1212–3.091 respectively. The peak positive weight was located in same place inside BA44 for both FBM models and the SVM model (figures 5.11 and 5.15 – middle and bottom rows). Adjusting the threshold to isolate the same number of voxels as 5% of the true volume results in the formation of a centrally placed cluster within BA39 and further expansion of the cluster in BA44 (figures 5.11 and 5.15 – middle and bottom rows).

Unlike the SVM model the 2 clusters continue to expand along the superior-inferior axis, with some extension beyond the Brodmann borders. There are no voxels identified within the STG until 40% (figures 5.14 and 5.16 – middle and bottom rows). The difference between the FBM 1000 iteration model and 2000 iteration model become clearer at 50%, with the former identifying a greater number of voxels located between the BA39 and BA44 (figures 5.14 and 5.16 – middle and bottom rows).

For each model the threshold is varied to select a volume of voxels equal to the some percentage of the combined BA39 and BA44 volume. All voxels identified by the model that lie within either BA39 or BA44 are treated as true positives, whilst those that lie outside are false positives. Thus at each percentage the sensitivity and specificity can be calculated to create a receiver operating curve (ROC). Figure 5.25 displays the ROC for the SVM, FBM 1000 iterations (FBM1000) and FBM 2000 iterations (FBM2000) models. Initially the SVM curve has the highest level of accuracy. The FBM1000 model eventually merges with the SVM curve at 56% of the true volume. Increasing the total number of iterations to 2000 results in the FBM2000 curve intersecting the SVM curve at 42% (145 voxels) of the true volume to lie above the SVM curve. Consequently the FBM2000 model correctly identifies more of the true volume than the SVM model overall.

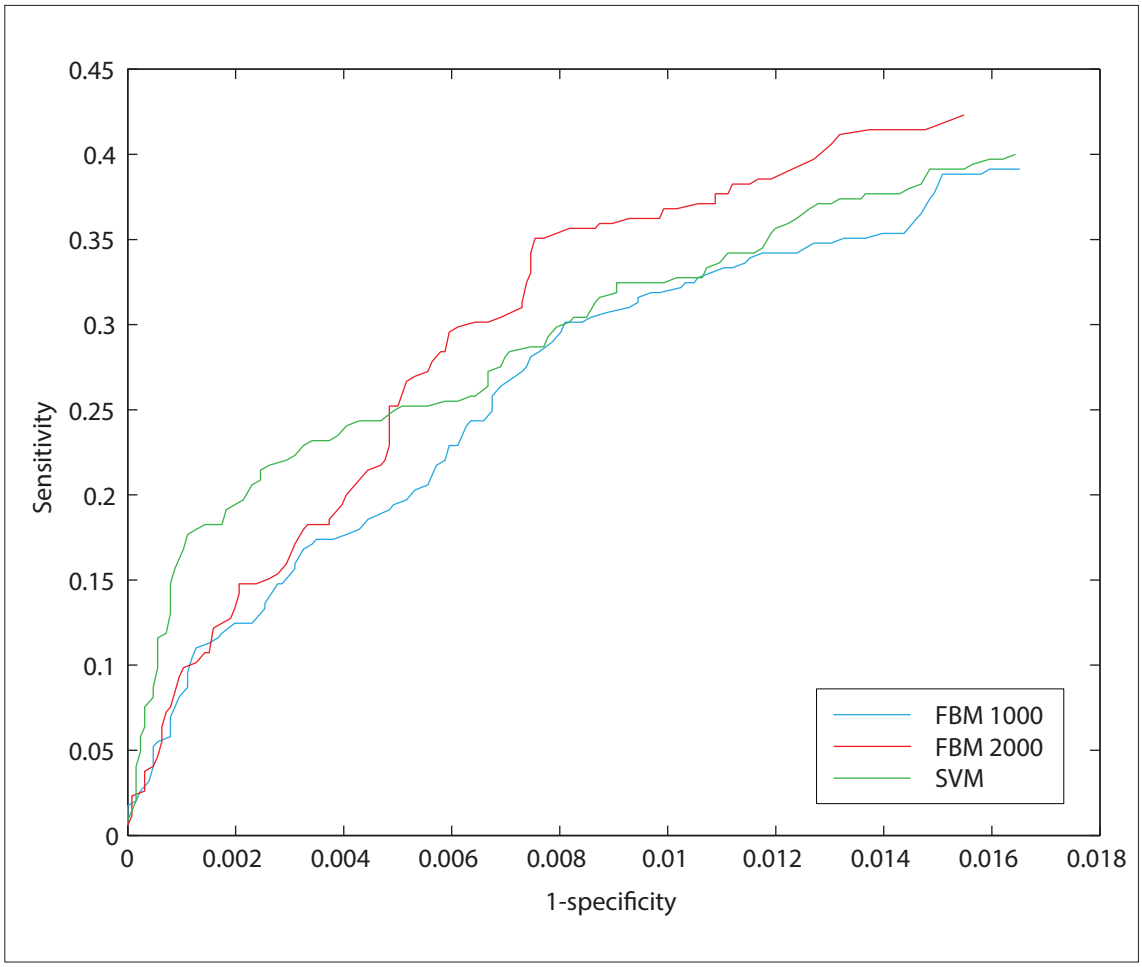


Figure 5.25 - Receiver operating curve comparing the SVM, FBM1000 and FBM2000 models.

The sensitivity and specificity for the SVM, FBM1000 and FBM2000 models is calculated for different fractions of the total “ground truth” volume specified by Brodmann area 39 and 44. The SVM curve intersects with the FBM2000 model at 42% and merges with the FBM1000 model around 56%.

5.3.2.4 Predictions

All 3 multivariate techniques achieved a high level of performance assessed by sensitivity, specificity, positive predictive value, negative predictive value and accuracy with respects to predicting whether a novel brain scan was "affected" or "unaffected" based on the original hypothetical model. The RVM model performed the worst out of the 4 models. The performance of the 2 FBM models was identical. Although there is no numerical difference in the performance between the linear kernel SVM model and FBM models, the scans classified incorrectly by each technique were different.

Dataset	Method	Sensitivity	Specificity	PPV	NPV	Accuracy
42	SVM	0.952	1.000	1.000	0.955	0.976
	RVM	0.905	1.000	1.000	0.913	0.952
	FBM1000	0.952	1.000	1.000	0.955	0.976
	FBM2000	0.952	1.000	1.000	0.955	0.976
156	SVM	0.952	1.000	1.000	0.993	0.994
	RVM	0.905	0.970	0.826	0.985	0.962
	FBM1000	0.952	1.000	1.000	0.993	0.994
	FBM2000	0.952	1.000	1.000	0.993	0.994

Figure 5.26 - Table showing the predictive performance of the 4 different multivariate models.

The 4 different multivariate models were tested on a novel test set of 156 scans. A balanced set of 21 affected and 21 unaffected scans is shown in the top half of the table, while the bottom half shows the performance on the entire test set. The SVM model is based on a linear kernel with a *C* parameter of 2^{-9}; the RVM model was trained using a maximum of 5000 iterations and a Bernoulli likelihood; and the 2 FBM models were generated using 1000 iterations and 2000 iteration with a burn in period of 500 and 1000 respectively. (PPV: positive predictive value, NPV: negative predictive value).

5.4 Discussion

5.4.1 Simulation one : Comparison of a mass univariate technique (Fisher's exact test) with a multivariate technique (SVM)

The first series of simulations examined a 2 loci model where the hypothetical functional deficit was dependent on damage to 20% or more of either Brodmann area 39 (BA39) or 44 (BA44) 90% of the time. The simulation reveals that in such a situation conventional lesion mapping would show an erroneous displacement of the inferred critical regions in a pattern not readily predictable from the single locus displacement field, to a region in a difference lobe, the superior temporal gyrus (figures 5.8 and 5.9).

The value at which to threshold the *p* value map to obtain the most informative representation of the lesion function relationship is unknown, and the decision to use a *p* value of 1% Bonferroni corrected for multiple comparisons is rather arbitrary. Alternatively we may examine the pattern of *p* values across the brain, with 3 distinct clusters gradually forming, one of which is located outside the 2 critical areas in the superior temporal gyrus (STG) (figure 5.4). Importantly, there is a tendency for the cluster in the STG to preferentially expand rather than to populate the frontal and occipital poles of Brodmann areas 44 and 39 respectively (figures 5.5 and 5.6). As a consequence the cluster in BA39 eventually merges with the one in the STG (figures 5.6 and 5.7). Although a conventional multivariate ANOVA may be considered to overcome the limitations of the Fisher's exact test to model the high dimensional multivariate distribution, this technique is not able to handle the dimensionality mismatch of the data as there are too many variables in proportion to the number of cases for this approach to be valid.

In relation to the same simulation the multivariate technique produces a weighting that indexes each voxel's contribution to the classification process. Gradually manipulating the weighting threshold results in 2 clusters that slowly expand to fill the critical regions, BA39 and BA44, without extraneous clusters appearing in distinctly separate regions of the brain. However the weights assigned to each dimension (voxel), cannot be used to infer criticality of any dimension, they are merely suggestive. Nevertheless by modelling the high dimensional spatial distribution with a Support Vector Machine (SVM), thresholding the brain map of weights to yield the same number of surviving voxels as the mass univariate approach showed a much better correspondence with the ground truth – compare the left (Fisher's exact test) and right (SVM) columns in figures 5.8 and 5.9. The fidelity of the SVM model is shown to be better than the mass univariate approach across the entire volume range of the critical region, with the ROC for the mass univariate (Fisher's exact test) technique lying below the SVM model.

The mass univariate approach has been shown to be susceptible to mislocalisations in the 2 loci scenario, with the displacement from the ground truth in excess of what is found in the single locus models resulting in a localisation in a different lobe to the critical regions. Critically, the vector of this displacement results in a shift unpredictable by the single locus models. The alternative of a multivariate technique offers a potential solution that is less perturbed by these hidden biases.

5.4.2 Simulation two : Comparison of 3 different multivariate techniques on a two loci model

The performance of the generated models from the 3 multivariate techniques can either be assessed with respects to their ability to infer the lesion-function

relationship or their ability to classify a novel case as either affected or unaffected.

The Relevance Vector Machine (RVM) model performed the worst out of the 3 techniques, with respects to identifying the critical regions, with no voxels within the critical regions identified as a positive relevance vector (RV). The Flexible Bayesian Modelling (FBM) models using either 1000 (FBM1000) or 2000 (FBM2000) iterations performed far better, with the difference between them and the Support Vector Machines model (SVM) being very slight. The receiver operating curve (ROC) assessing the accuracy of the models as a function of the proportion of the total critical volume (figure 5.25), shows the SVM model to have a higher accuracy initially. The FBM1000 model eventually merges with the SVM model around 56% of the total critical volume, however increasing the number of iterations from 1000 to 2000 results in an improvement in the ROC such that the FBM2000 curve crosses over the SVM model at 42%. Theoretically the estimates of the weights associated with each dimension in the FBM models should improve with each iteration and is indeed demonstrated in this simulation, it is likely that further iterations with the same data may improve the FBM model further such that the initial performance difference with the SVM model is removed or possibly reversed. The trace plots and autocorrelation functions do provide a reasonable indication of when stationarity begins to emerge to help guide the size of the burn in period. This is of particular importance since despite the theoretical improvement with greater iterations, due to the computational costs of a full Bayesian approach a limit must be exercised.

The ability to infer function from structure is related to the predictive performance of the model and this is also reflected in the results, with the RVM model performing the worst. However the accuracy of the model is still above 95%, which would suggest the model has over-fitted to the training data by

identifying additional associations between the critical regions and hypothetical functional deficit to assist its classification process. There is no difference between the predictive statistics of the SVM and both FBM models, although the incorrect classifications between the SVM and FBM models are different. Similarly, since there is a dimensionality mismatch there is undoubtedly some degree of over-fitting occurring in both the SVM and FBM models. However unlike the mass univariate approach, since the multivariate techniques are explicitly modelling the spatial interactions, increasing the sample size will reduce the dimensionality mismatch and decrease the extent of over-fitting. In contrast the mass univariate approach cannot remedy this mislocalisation error with larger sample sizes, as the problem lies with a systematic bias rather than noise.

5.5 Conclusion

The relationship between structure and function in the brain will depend on the interaction between the multivariate lesion distribution and the brain functional architecture; since the latter is unknown, and is precisely what we are using lesion mapping to establish, we need to model the former explicitly. This cannot be achieved with this type of data by using a mass univariate approach. Multivariate techniques offer a potential solution to the problems that affect traditional univariate analyses of lesion data. Although there is a significant time cost associated with these methods, this is likely to be less problematic with the continual improvements in computing power. On the other hand owing to the large number of variables it also requires much larger numbers of cases than is usual in the lesion-mapping literature: how large exactly we shall only know once the lesion distribution has been sufficiently well characterized. Never the less these models, once generated, produce high prediction accuracies, with an excellent level of sensitivity and specificity, taking only a couple of seconds to compute.

The multivariate technique of choice however is still unclear. Computational cost is still a pertinent restriction, with a full Bayesian technique requiring significantly more time than Support Vector Machines. Although there is a suggestion the full Bayesian technique will generate more accurate models this improvement is not reflected in the predictive performance. In both SVM and FBM approaches reducing the dimensionality mismatch will be beneficial, however since the computational time in SVM increases linearly with training dataset size, the advantage of SVM over FBM may be eroded.

6 Post-stroke outcome prediction using acute stroke imaging and high-dimensional multivariate algorithms

6.1 Introduction

In ancient Greece, Hippocratic medicine placed greater emphasis on prognosis than treatment. This was only in part due to the relative impotence of available treatments, leaving physicians with evaluation and prediction of disease progression as the most valuable skills they could offer to their patients. Although over the years there has been a shift towards diagnosis and treatment with the development of new and effective treatments, the ability to predict the course of a patient's disease is still of great importance. Not only are patients eager to know their prognosis with regards various outcomes such as life expectancy, morbidity and disability, so they may take appropriate action and plan for their future, but also clinicians and hospitals too. Accurate predictions regarding a patient’s future is essential in order to determine what treatments may be required in the immediate, short and long term to help individualise a management plan, but also assist service provision, particularly in an environment with so much emphasis on financial constraints.

Many different prognosticating tools are available in neurology, with the Oxford classification of stroke (OCS) one of the most popular. In the acute setting of stroke injury, the Oxfordshire classification of stroke uses clinical signs and symptoms to classify the damage into 1 of 4 approximate brain locations. These 4 classes have then been correlated to mortality scores and thus provides an approximate 1 year and 5 year risk of mortality for the patient

(Bamford et al., 1991). Two key features of the OCS are the relative simplicity of the system, and therefore ease of implementation, and the use of information that would otherwise be collected in a standard neurological examination. Neuro-imaging, in particular magnetic resonance imaging (MRI) with diffusion weighted sequences (DWI), is gradually becoming a routine request in stroke medicine particularly with the formation of hyper acute stroke units (HASU) in the UK. Many studies have investigated the correlation between neuro-imaging and clinical outcome scores such as the Barthel Index, National Institutes of Health Stroke Score (NIHSS) and Rankin score (RS), with early models examining the correlation between lesion volume and patient outcome (Baird et al., 1997; Barber et al., 1998; van Everdingen et al., 1998; Saunders et al., 1995). van Everdinger and colleagues found that patients with first ever stroke suffering from lesions less than 22ml in volume would have a good outcome with a sensitivity of 75% and a specificity of 100%. Later work utilising lesion volume determined from DWI MRI scans obtained within 48hrs of brain injury was found to correlate with Barthel scores – an ordinal score used to measure performance in activities of daily living (Mahoney and Barthel, 1965) – where individuals with 85 or more would be independent at discharge, whilst those below would be dependent (Thijs et al., 2000).

Interestingly median volume of injury in the right hemisphere is consistently larger than the left hemisphere for the same level of deficit measured by the NIHSS (Lyden et al., 2004; Woo et al., 1999), potentially suggesting the lesion pattern rather than absolute volume is of greater importance with regards creating a predictive tool. There have been a few studies that have tried to combine both location and volume into the same predictive model (Bang et al., 2005; Nazzal et al., 2009; Zhu et al., 2010). Many of these attempts to parameterise the pattern of injury have necessitated a significant amount of dimensionality reduction to facilitate the analyses. Reductions have included simplifying lesion volume to a set of coarse categories such as small, medium

and large; or partitioning the brain into a handful of locations such as lobes. Intuitively, assuming an appropriate sample size and technique are available, increasing the available information to a model should improve its predictive power. Therefore it is hypothesized that if we can maintain the spatial resolution and account for the interactions between dimensions (in this case between voxels), we should be able to improve upon the predictive power of current models.

In the above examples, the predictive models have attempted to correlate neuro-imaging obtained acutely in the admission with some outcome measure, generally one of patient functionality. The Barthel Index is an ordinal scale used to measure performance in activities of daily living based on a 10 item assessment. Although after an assessment a patient is given an integer score, many studies later binarize their outcomes into patients who are independent and dependent. The value of this threshold varies in the literature depending on the studies' definition of independence. Some groups have utilised a very strict cut-off with only individuals who obtain the maximum score being classed as independent (Johnston et al., 2000; Kwakkel et al., 2011; Uyttenboogaart et al., 2005), while others have specified scores of 60 and above out of 100 as an indicator of independence (Granger et al., 1979; Nakao et al., 2010). Work by Sulter et al rephrased the classification question and specified those who have a poor outcome as having a score less than 60 as there appears to be less disagreement in the literature with this threshold (Sulter et al., 1999). Following on from this, some groups have proposed using the Barthel Index (BI) as a means to prognosticate patients. Rollnik demonstrated that the BI obtained in the acute setting was predictive of a patient’s length of stay (Rollnik, 2009). Nakao examined the predictive potential of Barthel Index scores obtained within 21 days of stroke onset and found they were reliable predictors of the patients’ functional ability at 6 months post stroke injury (Nakao et al., 2010).

Previous clinical predictive tools have generally used either a single variable, univariate, or a handful of variables, “oligovariate”, as a way to model human physiology (Atlas, 2007; Fuchs and Smith, 2001; Holm et al., 2007; Milhorn et al., 1965). As a result significant effort has been invested in identifying the optimal variable reduction without questioning whether or not reduction is a viable avenue to pursue or not. In the case of the human brain, there is evidence to support the notion it is a highly complex distributed structure (Bullmore and Sporns, 2009; Catani and Thiebaut de Schotten, 2008; Fair et al., 2009; Greicius and Kimmel, 2012; Haxby et al., 2001; Mesulam, 1981). No reductive measure is ever going to work, *a priori*. Twenty to thirty years ago tackling this challenge, with minimal reductionism, using a multivariate approach would not have been possible, or at least without the help of a CRAY-2, “super-computer”, or two. In recent years, not only has the availability of computing power increased significantly but critically mathematical techniques, broadly referred to as machine learning, have appeared that can exploit this computing power and allow us to model very high dimensional relationships. In the following chapter I shall use the multivariate technique, Support Vector Machines (§1.3.3.2.1), and apply it to a clinical outcome measure as an example.

6.2 Methods

6.2.1 Hardware

All calculations were performed in the MATLAB environment (The MathsWorks Inc.).on an Intel i7 920 2.66GHz processor, with 6GB RAM and 120GB solid state drive (SSD).

6.2.2 Data

Diffusion weighted magnetic resonance imaging can identify the region of acute infarction within hours of symptom onset (van Everdingen et al., 1998; Warach et al., 1995, 1996), however, there is greater variability in the size of the delineated lesion with scans obtained less than 6 hours after symptom onset (Rivers et al., 2007). This variability is likely to be a disadvantage for the multivariate technique.

A total of 93 brain scans with paired Barthel Index (BI) assessments were collected from the National Hospital for Neurology and Neurosyrgery and the Korea Univeristy Anam Hospital. Figure 6.1 shows the lesion coverage map for the 93 scans.

6.2.2.1 National Hospital for Neurology and Neurosurgery images

A total of 75 brain scans were collected from the National Hospital for Neurology and Neurosurgery. All brain scans possessed a vascular lesion and were performed on a GE Genesis Signa 1.5 Tesla (1.5T) MRI scanner in a single session for each patient. The scans collected were the axially acquired T2-diffusion weighted echoplanar images (b0 and b1000 sequences) sampled at 1mm x 1mm x 6.5mm (TR=10000ms, TE=104.9ms, Columns=256, Rows=256,

Slice thickness=5mm). The mean delay between symptom onset and scan was 6.3 days (standard deviation (SD) 5.8 days). The mean age was 65.8 years (SD = 18.4 years), and the proportion of males was 0.45.

6.2.2.2 Korea University Anam Hospital images

The 18 images collected from Anam Hospital were all performed on a Siemens Magnetom, 3 Tesla (T3) TrioTim. The T2-weighted echoplanar images were harvested (TR=10000ms, TE=91ms, Columns=192, Rows=192, Slice thickness=5mm). The mean delay between symptom onset and scan was 0 days (SD 0 days). The mean age was 69.0 years (SD = 8.28 years), and the proportion of males was 0.43.

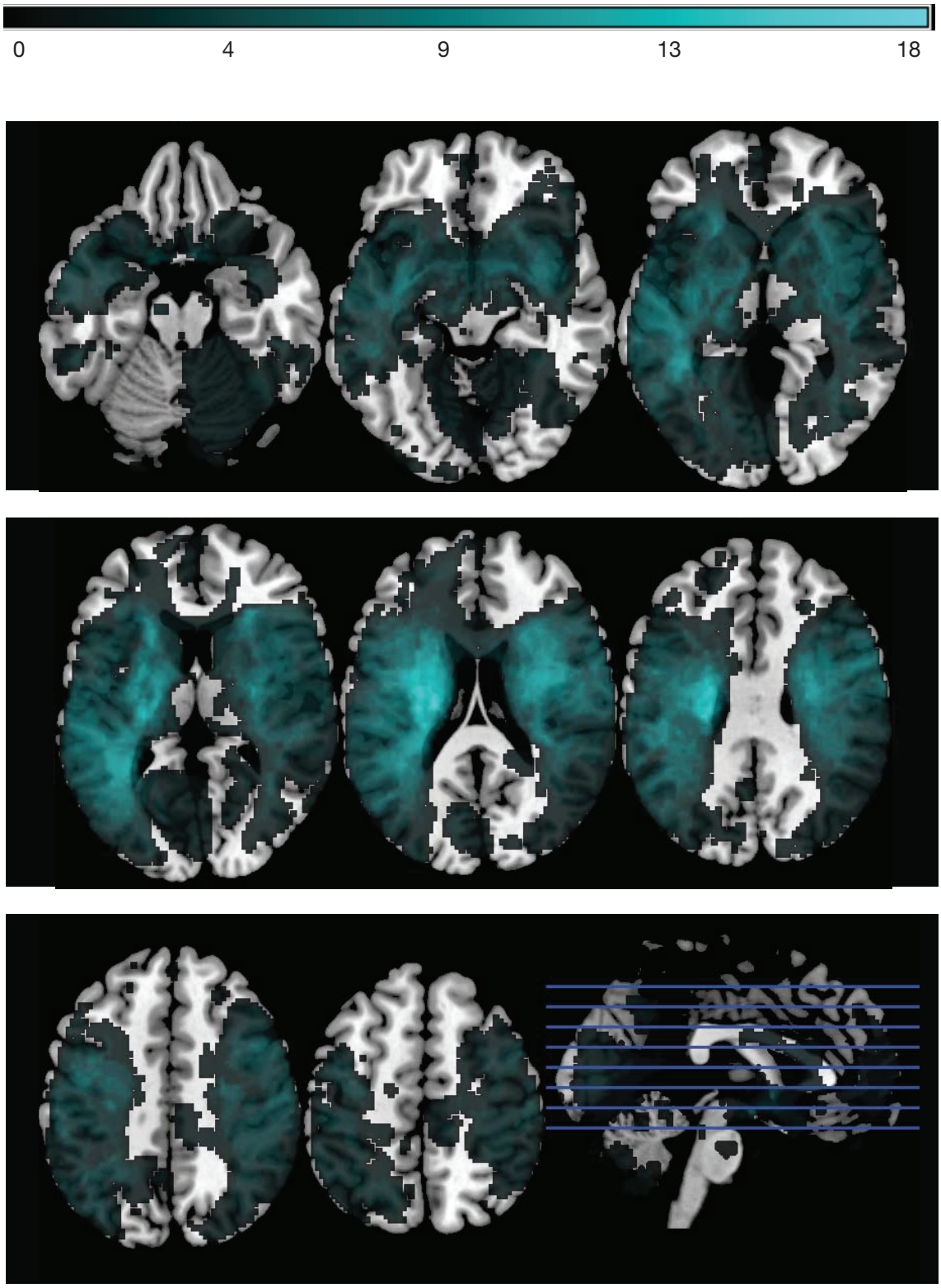


Figure 6.1 - Lesion coverage map.

The overlay (cyan) shows the coverage of the 93 segmented lesions. The axial slices are located at -20, -10, 0, 10, 20, 30, 40, and 50mm in MNI space. The peak overlay is 18.

6.2.2.3 Outcome data : Barthel Index (BI)

The paired outcome data consisted of a standard Barthel Index (BI) assessment. All scores were analysed on a scale from 0-100, with higher values representing a better functional performance. All BI were measured within 3 weeks of admission to hospital by a trained healthcare professional, with a mean delay period of 10.6 days (SD 7.20 days). The data was collected retrospectively from the patients' clinical medical notes.

6.2.3 Image preprocessing

For each patient the corresponding b0 and b1000 images were reconstructed into a 3 dimensional volume using xmedcon (Nolf et al., 2003). The b1000 volume was rigidly co-registered to the b0 volume using the co-registration function in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and its default parameter settings (the specific parameter settings are listed in appendix B). The b0 volume was subsequently spatially normalised into MNI space (ICBM space – European brains – template) using the combined segmentation-normalisation method in SPM8 (Crinion et al., 2007), with the derived normalisation parameters applied to the co-registered b1000 sequence. Again the default parameter settings for the segmentation-normalisation method provided by SPM were used, except for the interpolation setting in the writing option being set to a 6^{th} degree spline.

A binary mask for each patient was created via the lesion segmentation using the zeta anomaly detection method described in §2 (Mah et al., 2012). The resultant masks were sliced at 2mm isovoxels and occupied a volume measuring 91 x 109 x 91 voxels.

For the subsequent experiments only the within brain voxels were used. The brain mask template from SPM8 was thresholded at less than 0.8, and greater than 0.4 to identify the within brain voxels creating a 93 by 204234 data matrix.

6.2.4 Support Vector Machine (SVM) model generation

6.2.4.1 Classification of patients' outcomes into dependent and independent groups

Although the Barthel Index is a continuous scale, it is not uniform throughout its range. As discussed above, many studies binarize their measures into patients with good and poor functional outcome (Granger et al., 1979; Johnston et al., 2000; Kwakkel et al., 2011; Nakao et al., 2010; Sulter et al., 1999; Thijs et al., 2000; Uyttenboogaart et al., 2005). In line with other studies in the literature, the following sets of assessments were performed using 3 different thresholds. Barthel Index values for independence were specified as: 100; greater than or equal to 60 and greater than 40. Patients whose score were less than the threshold were classified as dependent.

6.2.4.2 Support Vector Machine model based on a linear kernel

In this scenario, each voxel is treated as a single variable that can possess one of two states; lesioned or unlesioned. Only voxels that were identified to be lesioned in at least 3 separate brain images were passed to the Support Vector Machine (SVM) algorithm. This reduced the total number of dimensions (voxels) to 25122.

To model the high dimensional relationships between the voxels the technique Support Vector Machines (SVM), was used. The software program LibSVM (Chang and Lin, 2001) was used to generate the model, based on a linear

kernel. Further details regarding the theory of Support Vector Machines and LibSVM can be found in §1.3.3.2.1.

A parameter search for the *C* value in the linear kernel was performed using a leave-2-out cross-validation technique. Each *C* value was calculated as 2^x, where the exponent *x* was an integer value between -20 and 20. A total of 41 different *C* values were assessed. For each value of *x* the dataset was split into a training set of 91 and a balanced test set of 2. The SVM algorithm was then presented the training data to generate a model based on a linear kernel. The test dataset, which has not been involved in the training process, is then used to assess the performance of the generated model. The test pair is therefore a novel dataset for that specific model. This process was iterated for every combination pair possible in the binarized dataset. In this way each trained model was naive to the corresponding test set pair, whilst minimizing the dimensionality mismatch.

The number of patients in either the independent or dependent group depended on the threshold Barthel Index value determining functional independence. These two sample sizes are not equal and it is important to account for the disparity as the standard notion of accuracy (the success rate of correctly classified examples) is not a good way to measure the success of a classifier applied to an unbalanced dataset since a model that merely predicts the majority class will perform well under it. This is because the algorithm assigns equal importance to the errors made on examples belonging to the majority and minority class. It is possible to assign a weighting parameter in LibSVM to adjust the cost for misclassification and account for the sample size disparity. The weighting adjustment is the ratio in sample size (larger group : smaller group). The nearest integer value of this ratio was used.

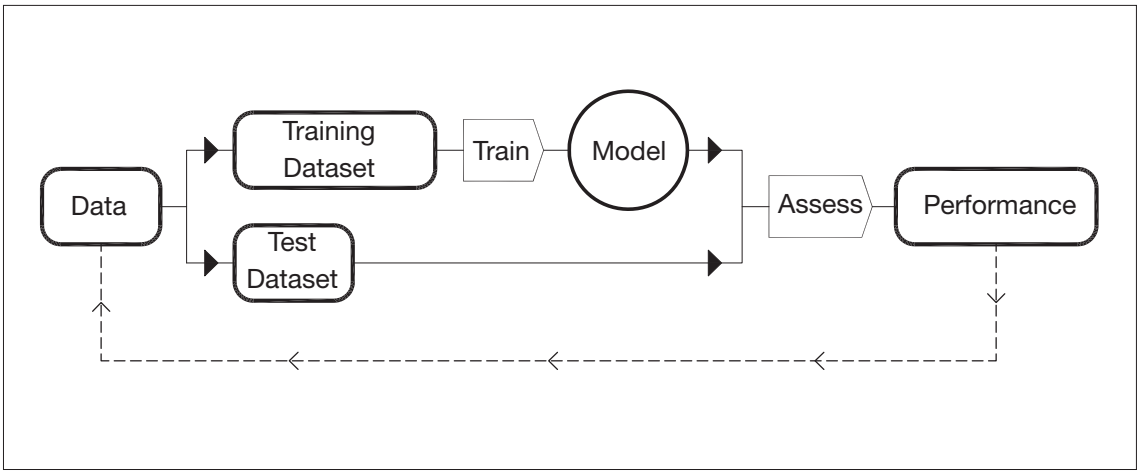


Figure 6.2 - Cross-validation flow diagram.

The available dataset of 93 patients is first divided into 2 sub datasets; training and test. In a leave-*n*-out cross-validation, the size of the test set is equal to *n*. In this case *n* is equal to 2. The training dataset is then presented to the SVM algorithm to generate a model based on a linear kernel. The predictive performance of the generated model is then assessed against the test dataset which has not been involved in the training process. The entire process is then repeated for a different training-test split.

6.2.5 Evaluation

The performance of the model was evaluated using the sensitivity, specificity, positive predictive value, negative predictive value and accuracy. Positive outcomes were those where patients were classified as independent by the Barthel Index score, while negative outcomes corresponded to scores reflecting dependence.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$positive\ predicitive\ value = \frac{TP}{TP + FP}$$

$$negative\ predictive\ value = \frac{TN}{TN + FN}$$

True positive (TP) = correctly identified as independent

True negative (TN) = correctly identified as dependent

False positive (FP) = incorrectly identified as independent

False negative (FN) = incorrectly identified as dependent

Since every test pair consists of a independent and dependent patient, the baseline accuracy of the model will be 50%.

6.2.5.1 Visualisation of the weights associated with each dimension

The hyperplane that represents the decision boundary within the *n*-dimensional feature space is defined as:

$$w \cdot \varphi(x) + b = 0$$

In this way, data points are classified according to which side of the plane they lie depending on whether the result is greater or less than zero. This transformation is described in the variable *w* which is a vector whose composite elements are the corresponding weights for each dimension in the model. Each weight will have a polarity and magnitude. If the polarity is positive then the associated dimension is important for the classification of the positive class, while a negative polarity is important for the other class. The magnitude represents the relative importance of the dimension (in relation to the other dimensions) to the classification process. Therefore dimensions whose weights have larger absolute values are more influential in the classification process than those with smaller values.

The weights associated with each dimension – in this case each voxel – were extracted from the optimal model trained on the entire dataset of 93 patients. Since the definition of functional dependence is specified as a negative outcome, this class is associated with voxels that possess weights that are less than zero. All negative weights generated by the linear kernel were mapped into their corresponding locations within the 3 dimensional brain volume. This generated a map of the regions of the brain that are considered influential to the expression of functional dependence. The anatomical labels of the areas affected were obtained from the John Hopkins University brain atlas provided in MRIcron (Rorden and Brett, 2001).

6.3 Results

6.3.1 Functional independence described as a Barthel Index of 100

Using a SVM model based on a linear kernel the parameter search for the optimal *C* value took 3.5 hours. The model with the highest accuracy rate was achieved with a *C* value of 2^{-3} (accuracy = 64.6%). The corresponding sensitivity, specificity, positive predictive value (ppv) and negative predictive value (npv) were 0.803, 0.488, 0.661 and 0.748 respectively. The accuracy deteriorates and then plateaus after 2^{2} (figures 6.3 and 6.4).

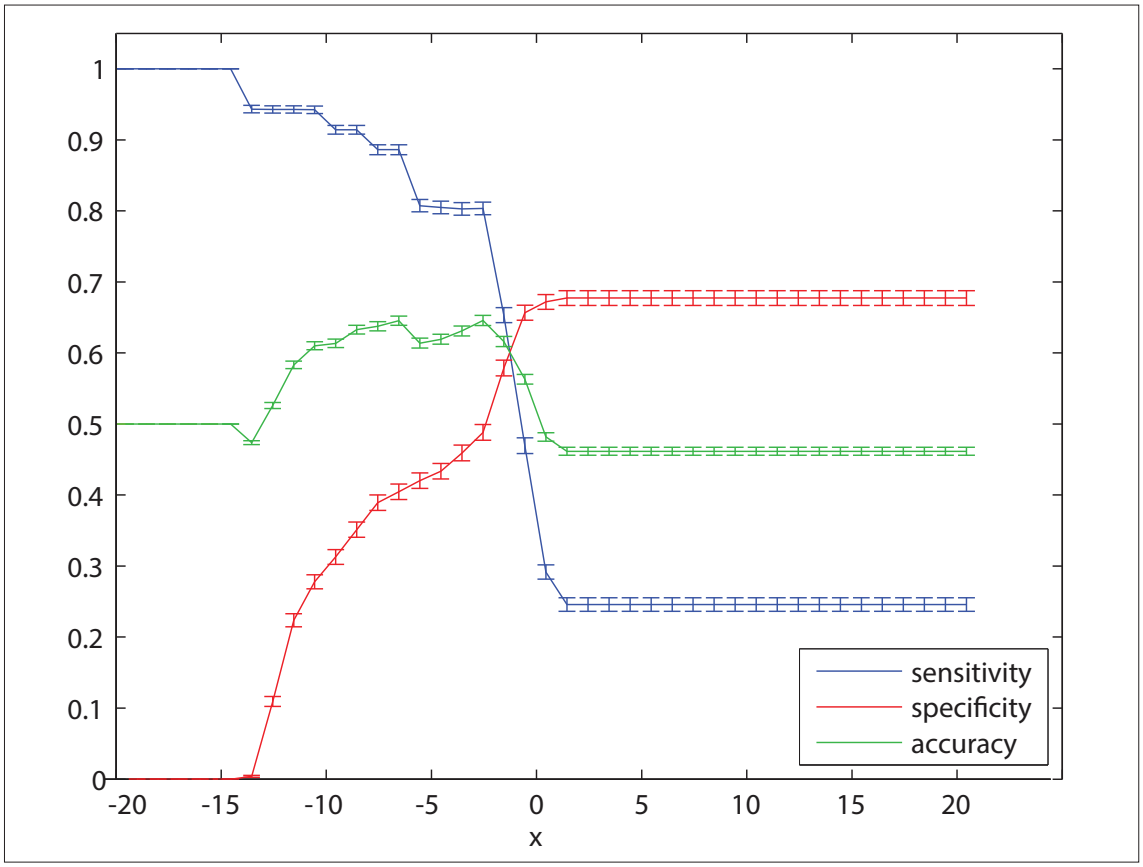


Figure 6.3 - Plots showing sensitivity, specificity and accuracy as a function of the *C* parameter for an SVM model based on a linear kernel: independence specified as a BI equal to 100.

Plot displaying the mean sensitivity, specificity and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index equal to 100, was performed. The peak accuracy was obtained with a *C* value of 2^{-3}.

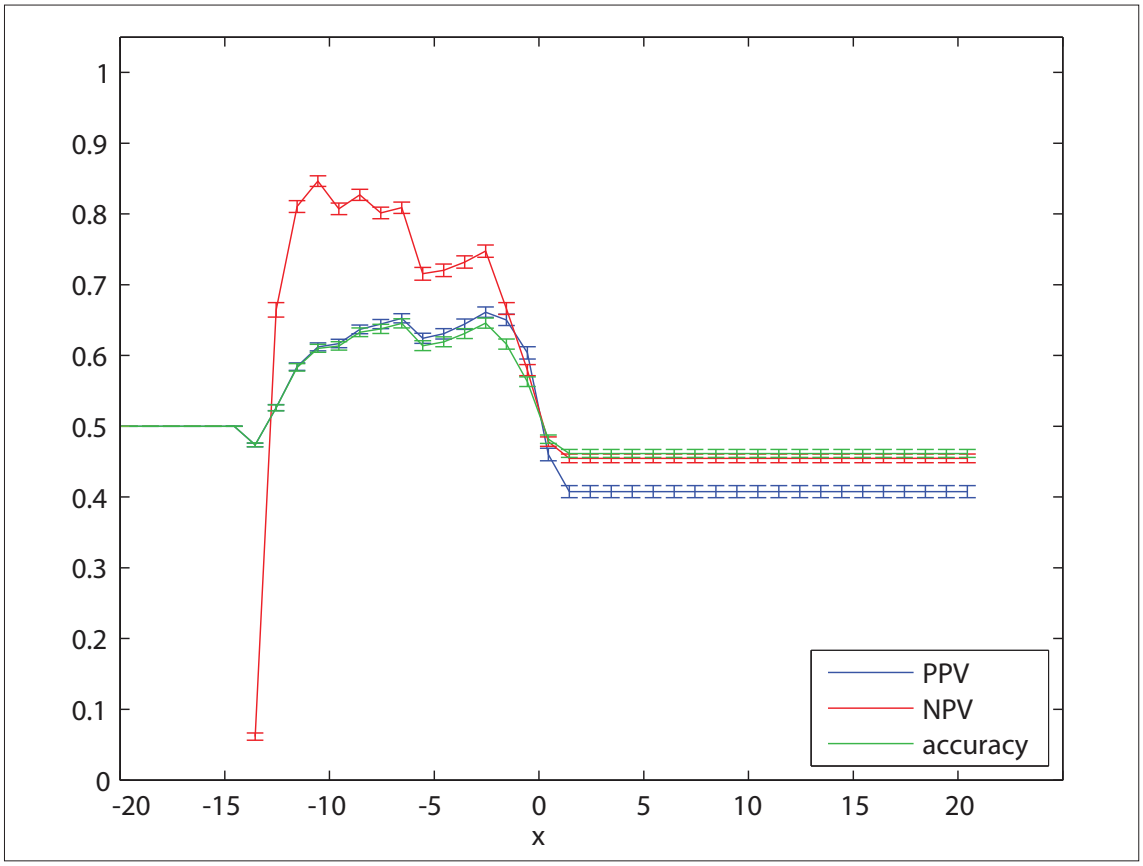


Figure 6.4 - Plots showing positive predictive value, negative predictive value and accuracy as a function of the C parameter for an SVM model based on a linear kernel: independence specified as a BI equal to 100.

Plot displaying the mean positive predictive value (ppv), negative predictive value (npv) and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index equal to 100, was performed. The peak accuracy was obtained with a *C* value of 2^{-3}.

6.3.2 Functional independence described as a Barthel Index of greater than or equal to 60

Using a SVM model based on a linear kernel the parameter search for the optimal *C* value took 2.5 hours. The model with the highest accuracy rate was achieved with a *C* value of 2^{-10} (accuracy = 71.2). The corresponding sensitivity, specificity, positive predictive value (ppv) and negative predictive value (npv) were 0.950, 0.484, 0.722, and 0.925 respectively.

The value of the *C* parameter reflects the trade off between the errors of the Support Vector Machine on the training data and margin maximization such that as *C* tends towards infinity the soft margin (§1.3.3.2.1.3) algorithm tends towards a hard margin (Rychetsky, 2001).

The accuracy of the model deteriorates slightly after 2^{-10}, and plateaus after 2^{-5} (figures 6.5 and 6.6). This would suggest that the 2 groups within the data – independent and dependent – classified according to this rule is more readily seperable, with a linear model, in the feature space compared with the split where functional independence is measured at a BI score of 100.

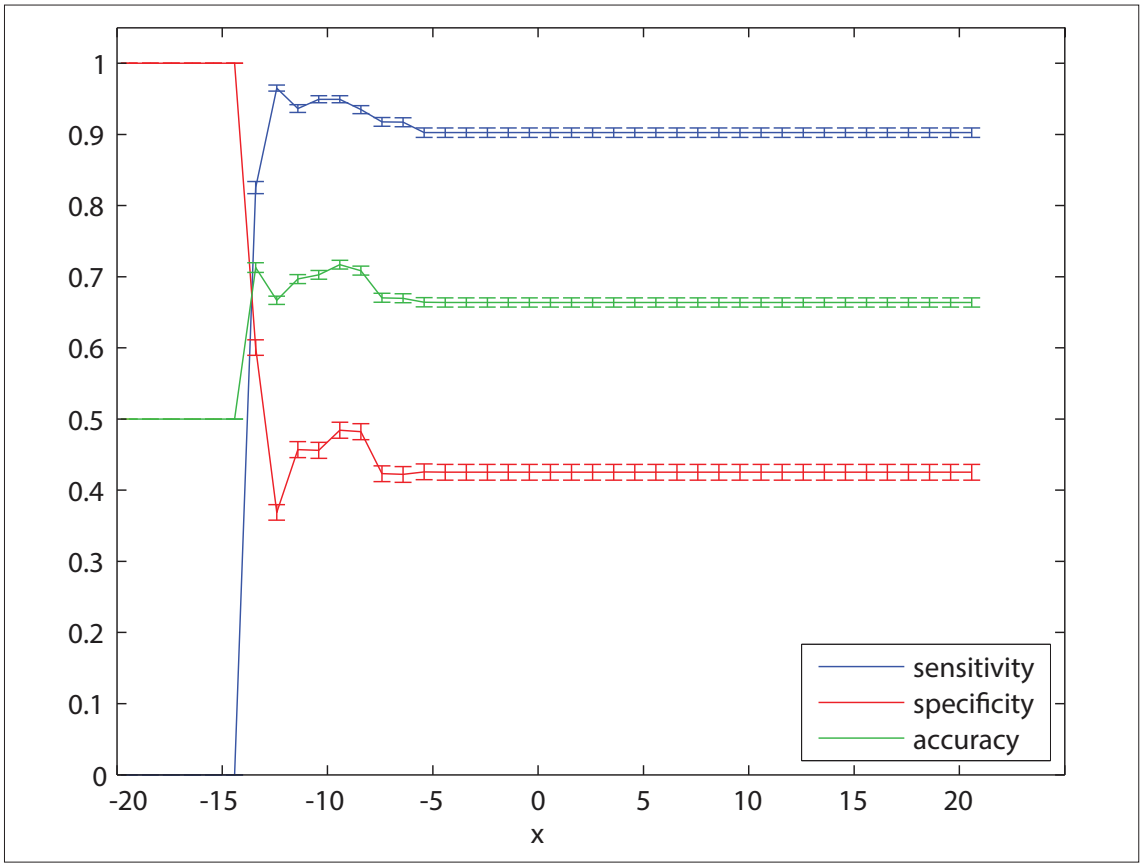


Figure 6.5 - Plots showing sensitivity, specificity and accuracy as a function of the C parameter for an SVM model based on a linear kernel: independence specified as a BI greater than or equal to 60.

Plot displaying the mean sensitivity, specificity and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index greater than or equal to 60, was performed. The peak accuracy was obtained with a *C* value of 2^{-10}.

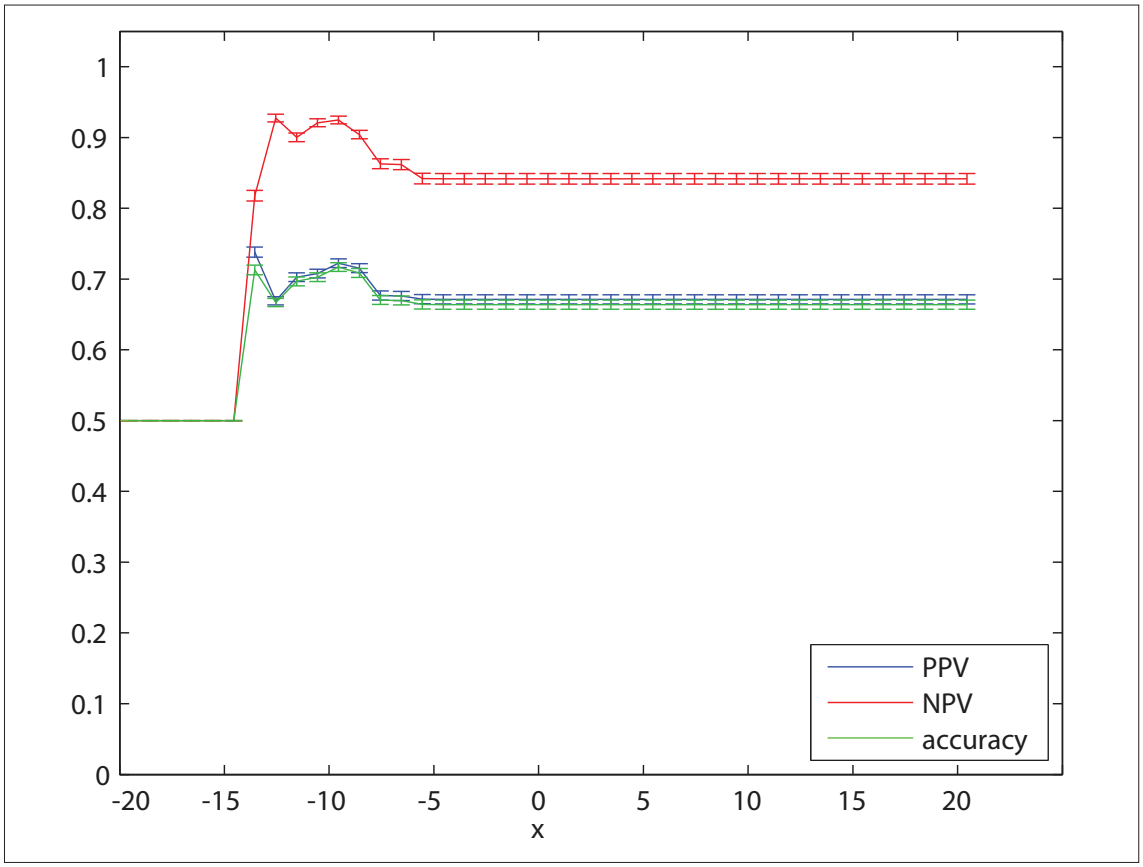


Figure 6.6 - Plots showing positive predictive value, negative predictive value and accuracy as a function of the C parameter for an SVM model based on a linear kernel: independence specified as a BI greater than or equal to 60.

Plot displaying the mean positive predictive value (ppv), negative predictive value (npv) and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index greater than or equal to 60, was performed. The peak accuracy was obtained with a *C* value of 2^{-10}.

6.3.3 Functional independence described as a Barthel Index of greater than 40

Using a SVM model based on a linear kernel the parameter search for the optimal *C* value took 2.4 hours. The model with the highest accuracy rate was achieved with a *C* value of 2^{-9} (accuracy = 75.6). The corresponding sensitivity, specificity, positive predictive value (ppv) and negative predictive value (npv) were 0.955, 0.557, 0.763 and 0.945 respectively.

There is little variability of the accuracy of the model with *C* values greater than 2^{-9} (figures 6.7 and 6.8). This would suggest that the 2 different classes – independent and dependent – overlap less in the high dimensional feature space compared with the model where functional independence is measured at a BI score of 100. This is also reflected with the optimal *C* value achieving a higher accuracy.

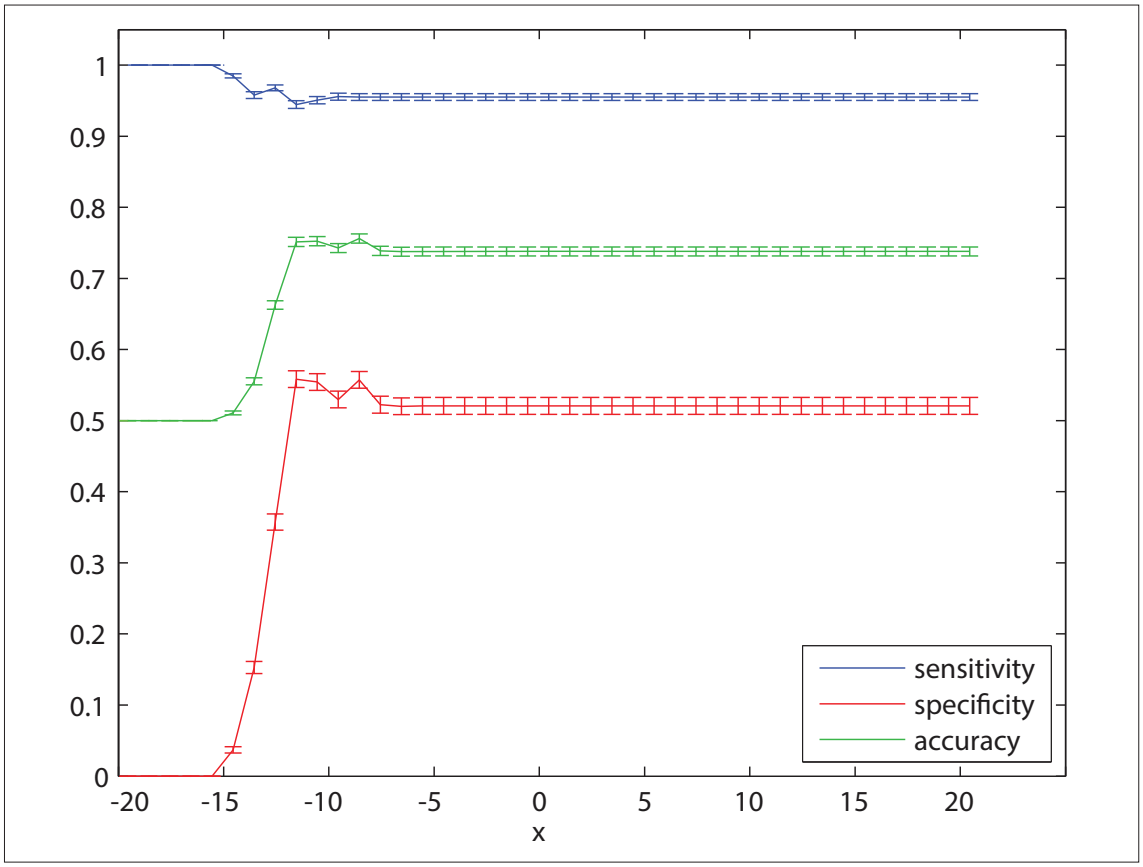


Figure 6.7 - Plots showing sensitivity, specificity and accuracy as a function of the *C* parameter for an SVM model based on a linear kernel: independence specified as a BI greater than 40.

Plot displaying the mean sensitivity, specificity and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index greater than 40, was performed. The peak accuracy was obtained with a *C* value of 2^{-9}.

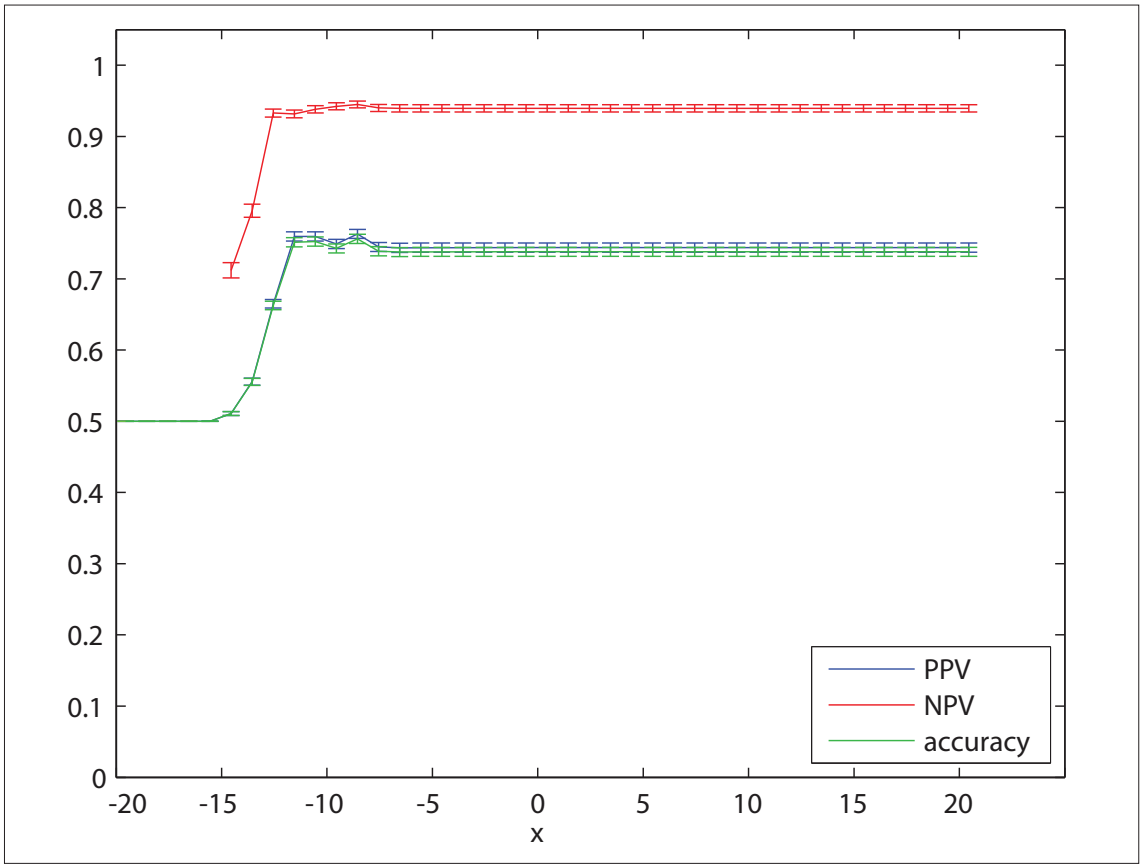


Figure 6.8 - Plots showing positive predictive value, negative predictive value and accuracy as a function of the *C* parameter for an SVM model based on a linear kernel: independence specified as a BI greater 40.

The mean positive predictive value (ppv), negative predictive value (pnv) and accuracy for different Support Vector Machine models based on a linear kernel with a *C* parameter of 2^x. The error bars are 2 standard errors of the mean. For each parameter setting a leave 2 out cross validation assessment of the 93 patient dataset, with independence specified as a Barthel Index greater than 40, was performed. The peak accuracy was obtained with a *C* value of 2^{-9}.

6.3.4 Visualisation of the weights extracted from the Support Vector Machine model

In all 3 models there is a greater representation of the left hemisphere by the negative weights. Weights with a larger modulus represent dimensions (voxels) that are of greater importance, relative to the other dimensions in the model, in the classification process. In this case for determining whether a patient is dependent. As the threshold value is lowered the weights with the largest modulus cluster around the left corona radiata and to a lesser extent on the right (figure 6.9 - 6.11).

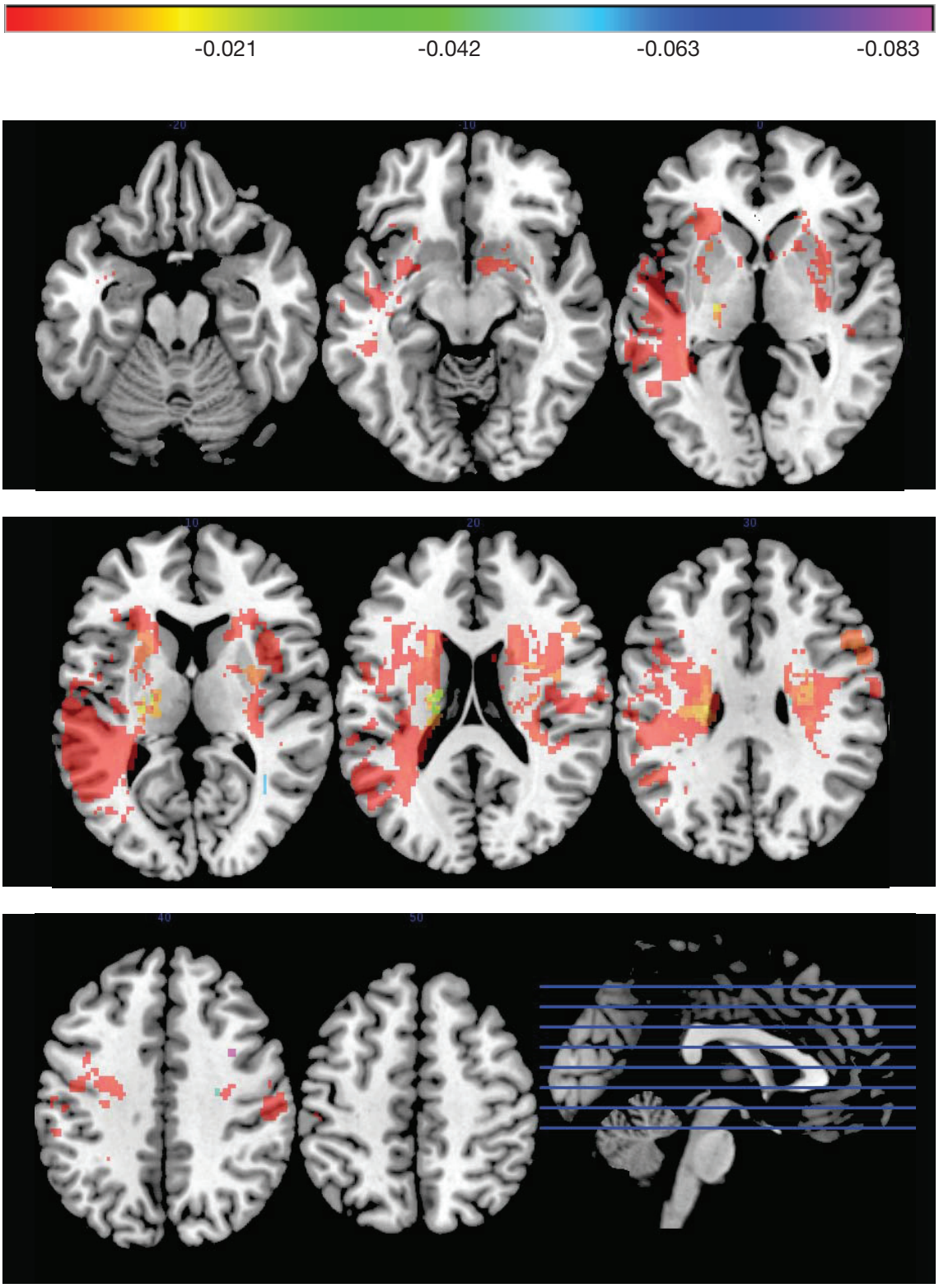


Figure 6.9 - Colour map of SVM model: functional independence described as a Barthel Index equal to 100.

Colour map of the negative weights associated with each dimension (voxel) extracted from the linear kernel SVM model trained on the entire dataset of 93 patients. A patient is classed as independent if their BI is equal to 100. Axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space.

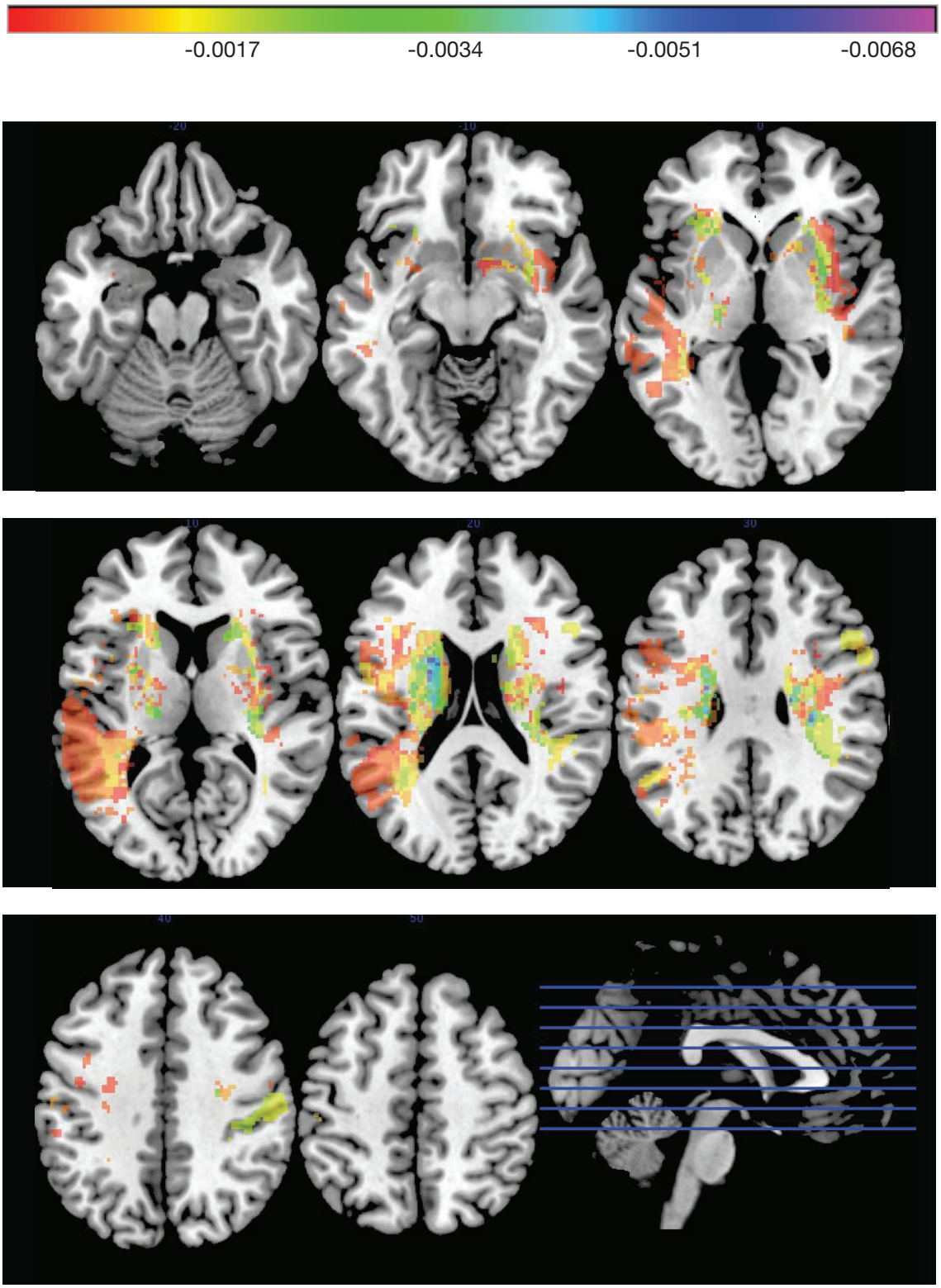


Figure 6.10 - Colour map of SVM model: functional independence described as a Barthel Index greater than or equal to 60.

Colour map of the negative weights associated with each dimension (voxel) extracted from the linear kernel SVM model trained on the entire dataset of 93 patients. A patient is independent if their BI is greater than or equal to 60. Axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. There is a cluster with the peak modulus weight in the left corona radiata.

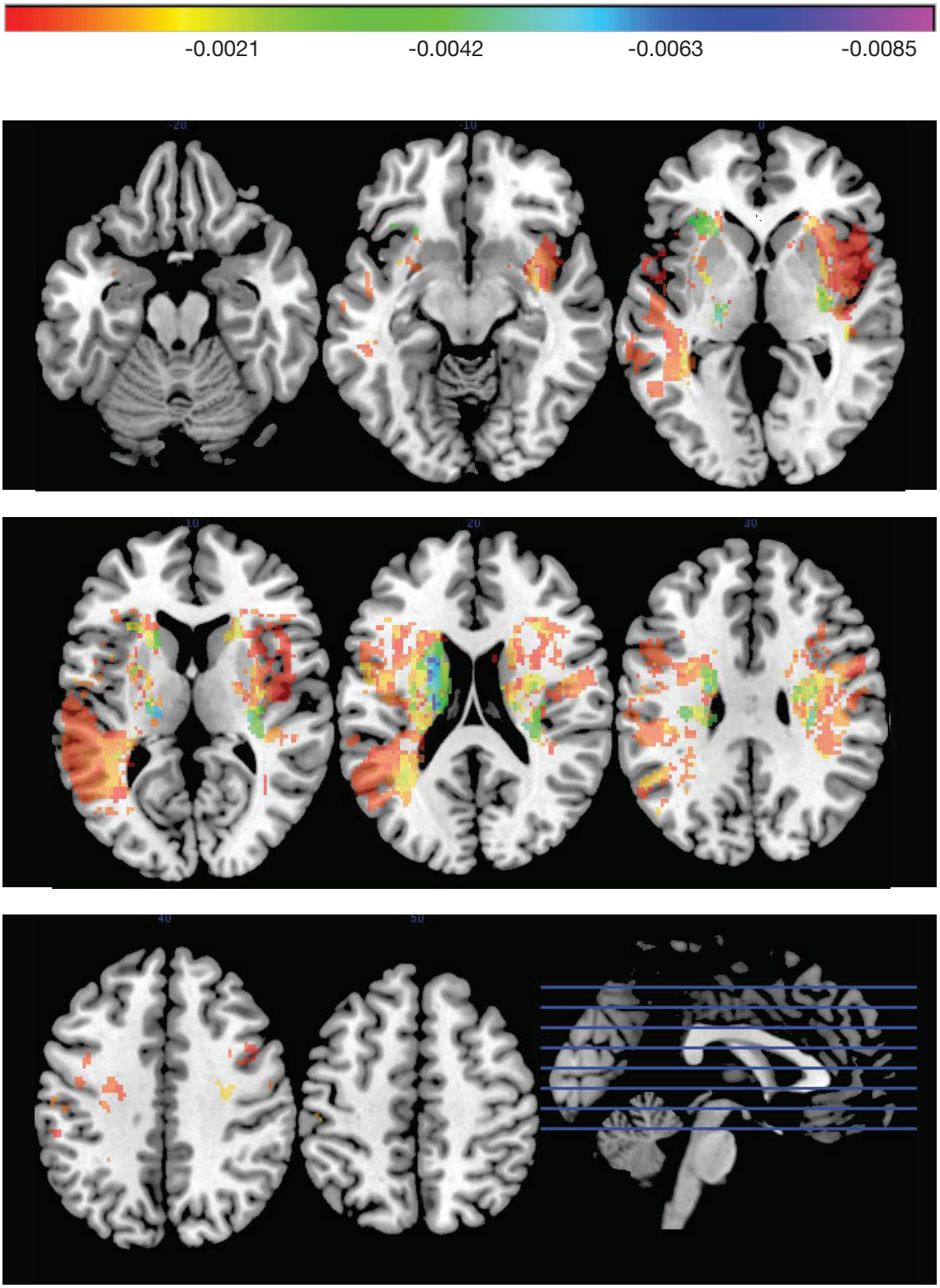


Figure 6.11 - Colour map of SVM model: Functional independence described as a Barthel Index greater than 40.

Colour map of the negative weights associated with each dimension (voxel) extracted from the linear kernel SVM model trained on the entire dataset of 93 patients. A patient is independent if their BI is greater than 40. Axial slices are located at -20, -10, 0, 10, 20, 30, 40 and 50mm in MNI space. There is a cluster with a peak modulus weight in both the left and right corona radiata.

6.4 Discussion

The experiments described here have demonstrated the application of the high dimensional multivariate technique, Support Vector Machine (SVM), to a clinical scenario. The models are able to predict whether a patient is dependent or independent with an accuracy between 65-75% based solely on the diffusion weighted magnetic resonance imaging performed in the acute setting of stroke.

The results of this study support previous work showing the potential of diffusion weighted (DWI) magnetic resonance imaging (MRI) in the acute phase of stroke as a predictor of future clinical outcome (Johnston et al., 2007; Kwakkel et al., 2011; Lövblad et al., 1997; Schiemanck et al., 2006; Thijs et al., 2000; Tong et al., 1998). It has been proposed that parameterising the lesion in neuro-imaging can improve the predictive power of such models, with a number of studies demonstrating this to be the case albeit with relatively low resolution compared with here (Bang et al., 2005; Schiemanck et al., 2006; Stone et al., 2000). Although our models are above chance, their predictive performances do not supersede existing ones (Johnston et al., 2007; König et al., 2008; Schiemanck et al., 2006; Stone et al., 2000).

Reasons for this finding may be due to the following. First, our dataset consisted of 93 patients with paired imaging and outcome data. Although this sample size would be reasonable with a univariate analysis, this presents a significant dimensionality mismatch between the number of data points and the number of dimensions employed with the high dimensional multivariate approach used here. To reduce this imbalance dimensions (voxels) with insufficient sampling were excluded, with only voxels implicated in at least 4 different lesions being passed on to the Support Vector Machine algorithm. As a result a comprehensive description of the entire brain volume was not possible thereby limiting the potential advantage of the spatial parameterisation

of the approach. It is reasonable to believe that increasing the sample size, a task that is not insurmountable, will result in improvements in the predictive performance as the description of the general population improves. Following on from this, the optimisation method of leave-2-out cross-validation was chosen as the most appropriate technique considering the size of the dataset. In general training on a 90% of the dataset and testing on the remainder in a 10 fold cross-validation arrangement is believed to be the optimal approach (Efron, 1983; Özsu, 2009). A smaller test set was utilised to minimize the dimensionality mismatch, with every balanced combination pair assessed to ensure both outcomes were equally tested. Selecting equal proportions for the independent and dependent groups would probably excessively handicap the algorithm.

Second, the Barthel Index (BI) is a summary measure of patients' functionality with respects to activities of daily living. Although the BI is a useful assessment tool with ease of administration and low inter rater variability, it is not very sensitive particularly in patients with mild stroke (Duncan et al., 1997). The summary nature of the BI may in fact complicate the situation with a variety of specific functional deficits secondary to a range of lesion patterns causing a similar change in the final score. Interestingly, the BI has a bias towards monitoring motor deficit with 8 out of the 10 questions focusing on complex motor tasks. When the weights, associated with each dimension, by the linear kernel model were mapped back on to the brain, the cluster of greatest importance attributed by the algorithm was located in the corona radiata an area identified as predictive of future motor function after stroke (Chen et al., 2000; Cho et al., 2007; Shelton and Reding, 2001). Similarly, a prediction model will be limited by the variance inherent in the outcome measure. Although an early Barthel Index has been shown to be representative of the level of independence at 6 months (Liu et al., 2004; Nakao et al., 2010), measures very early on in a patient's admission will generally underestimate the BI. In the

dataset used here 18 patients had a BI measured within 24 hours of admission. Using more targeted outcome measures, such as the National Institute of Health Stroke Score (NIHSS), and / or less inherent variability may result in more accurate models and improved prediction accuracies.

Third, many of the existing models have used information collected from the patient's clinical history and functional performance scores (Johnston et al., 2000; König et al., 2008; Schiemanck et al., 2006). This will undoubtedly provide influential information, such as co-morbidities, regarding the patient's prognosis. Although it is possible to incorporate these features into our model, obtaining this information can be time consuming and difficult particularly in the acute stages therefore limiting the utility of the predictive tool. However relying solely on the DWI scan, an investigation that is becoming standard protocol in stroke medicine, simplifies the implementation of the tool while not interfering with the patient's acute management. If the prediction accuracies are comparable to those that require non-radiological information despite the former limitations, this limitation can either be easily rectified by the inclusion of these features, or may not be necessary with the correction of the limitations discussed earlier.

6.5 Conclusion

The foregoing experiments have shown the potential of spatially parameterising diffusion weighted magnetic resonance imaging (MRI) taken in the acute phase of stroke management in generating predictive outcome models. Increasing the available training data and utilising more specific outcomes will further improve their predictive performance which may be useful in stratifying patients in research trials as well as guiding clinical management. With the use of diffusion weighted MRI scans routinely taken in the management of stroke becoming more widespread, the implementation of such a tool is relatively simple and unobtrusive to current clinical practice.

7 Conclusion

It can be argued that the principal objective of modelling the human body is to predict how it will respond in various physiological and pathological states. Ideally this is achieved by constructing a model that emulates as closely as possible the same structure as the underlying true system. With respect to human brain mapping this involves trying to construct a model that explains which regions of the brain are involved in carrying out specific functions. The anatomically specialised organisation of the human brain makes this a possibility.

With both functional magnetic resonance imaging (fMRI) and lesion mapping the controlled variable is the behaviour or putative function, either through manipulation of the experimental task or by separating the population into those with and without a deficit or other functional feature of interest. The observed variable are the areas of the brain affected. Functional MRI has been very useful in identifying regions within the brain that are active during particular tasks, and has thus guided the generation of hypotheses that implicate specific brain regions in specific functions. However, in order to determine whether or not a region is critical for a specified function we need to be able to show that in its absence the function fails. Lesion mapping studies provide a framework for testing these hypotheses with potentially stronger inferential power than correlative techniques such as fMRI. But the framework is not a simple one – conceptually and methodologically.

7.1 Spatial biases within lesion imaging data

Past lesion studies have employed techniques where each unit volume within the brain, such as a voxel, is treated as spatially independent from the next. These include techniques such as the template overlay method and more recently with voxel based lesion symptom mapping (VLSM) (Bates et al., 2003).

The issue of spatial independence was explored in §4 focusing primarily on the technique of VLSM, where each voxel within the brain is treated as a spatially independent variable resulting in a mass univariate approach to modelling the brain. Our concerns arise from the possibility, indeed likelihood, of hidden biases in the structure of lesion data which are unconnected to the function we are examining yet distort the inference. If the inter-voxel dependencies are not accounted for then the mass univariate method may be susceptible to displacement secondary to these hidden biases, resulting in mislocalisation of the structure-function inference.

In the first series of simulations, a cohort of 581vascular injured brains were spatially parameterised, with each voxel representing a separate dimension whose state was either healthy or lesioned. This high dimensional dataset was then embedded into a 2 dimensional feature space where the location of each brain volume is such that the high dimensional relationships between the different brain volumes are preserved. Critically the resultant images (figure 4.5 and 4.6) do not show a uniform distribution in the 2 dimensional feature space, but a series of clusters indicating that the lesion dataset possess some form of spatial structure. The voxel-wise simulations (§4.2.3.2.1) and Brodmann area simulations (§4.2.3.2.2) that examined the dependence of a putative function of interest on either a single voxel or cluster of voxels respectively quantified the mean displacement from the true location in the order of 15.7mm. This

magnitude of displacement is sufficient to displace our localisation across Brodmann area boundaries and even lobes.

Although the direction of displacement from the true locus has been illustrated in the form of a vector, these displacement field maps (figure 4.7) are specific to the model we explicitly defined prior to the simulation. Consequently these displacement field maps cannot be used in an *ad hoc* manner to correct for these biases as the map is dependent on the underlying model we are trying to identify.

Indeed, as we demonstrate in §5 taking Brodmann areas (BA) 39 and 44 as an example, to model the kind of anatomically distributed system likely to obtain in reality requires a wholesale change of approach to high-dimensional multivariate inference. The mass univariate technique displayed a tendency to identify areas outside of the critical regions in preference to a large proportion of BA39 or BA44 (figures 5.8 and 5.9 – left column). Also the centre of the significantly associated region did not fall in either Brodmann areas. Importantly the mislocalisation displayed by the mass univariate approach is not readily predictable from the single locus displacement field maps shown in figure 4.7.

We propose that a potential solution to this problem is the use of a multivariate technique that will account for the spatial dependencies within the data. Using the same data set as was used with the mass univariate approach, the two loci simulation was repeated using the multivariate technique Support Vector Machines (SVM). When comparing the areas identified by the technique both BA39 and BA44 are populated preferentially to those outside of the critical region. As a result the SVM model shows a much better correspondence with the ground truth compared with the univariate technique and would appear to be less susceptible to the spatial biases present within the data. Importantly the mislocalisation displayed by the univariate technique is not secondary to

noise which would be ameliorated by an increase in the sample size but to the spatial biases within the data. Indeed the current simulation uses a cohort of 581 patients which is one of the largest sample sizes used in a lesion study. However it is probable that an increase in sample size will further improve the multivariate model as this will reduce the dimensionality mismatch and assist the algorithm in handling the spatial dependencies between dimensions.

Unlike the mass univariate technique which assigns a *p*-value to each dimension (voxel) the SVM model assigns a weight. This therefore only provides an indication of the relative importance of one dimension with respects to another; it cannot be used to infer criticality of a dimension. Similarly it is not clear how the brain map of weights should be thresholded. With the univariate approach a *p*-value of 5% (corrected for multiple comparisons) is often proposed however this in itself is also rather arbitrary. In contrast with this, all three multivariate techniques (Support Vector Machines, Relevance Vector Machines and Flexible Bayesian Modelling) generate weights that possess a magnitude and polarity. Therefore it is possible to state a rather coarse threshold at zero which implies the dimension is neither predictive for or against the specified outcome.

The relationship between structure and function in the brain will depend on the interaction between the multivariate lesion distribution and the brain functional architecture. Since the latter is unknown and is precisely what we wish to establish, we must therefore model the former explicitly. This cannot be achieved with this type of data using a mass univariate approach. The multivariate techniques described offer a potential solution to this problem, but have in the past been faced with obstacles such as limited computational power and small sample sizes. It has been shown that the multivariate technique, SVM, is better adapted to handle the spatial biases within the data with a relatively small sample size for this type of analysis (the dimensionality

mismatch was 581 vs 90469) within a reasonable time window. The computational limitations are a minor issue in this case, as these calculations have been performed on commercially available computer workstations. The time may be reduced by improvements in processor clock speeds and parallelisation of the software. The solution to the dimensionality mismatch is simply to increase the available data. This is not an easy task if the number of dimensions is in excess of tens of thousands. The feasibility of these studies will increase with greater automation: this is the reason for focusing here on automated image analysis techniques.

7.2 Automation of image analysis

In lesion-based brain mapping studies, the brain images must be first undergo lesion segmentation and spatial normalisation before any meaningful statistics can be performed. Although automation of these steps will facilitate the creation of larger samples, it can also help minimise the effect of noise, either simply through the larger dataset, or by reducing the intra and inter operator variation. Crucially, removing operator dependence reduces the risk of unpredictable bias.

In §2, we proposed an unsupervised lesion segmentation algorithm and assessed its performance against 38 manually segmented brains with vascular lesions. The algorithm was shown to closely emulate manual segmentation with a sensitivity and specificity of 0.960 and 0998 respectively. There was very little variability across both subjects and lesions within which bias could conceivably emerge, and indeed less than what would typically be found between observers. Although it is not possible to assert that the algorithm is better than manual segmentation, it may however be more consistent. Minimizing the amount of variability within which bias may emerge is crucial,

since as the sample size increases, although the effect of noise will decrease, bias on the other hand will persist and may in fact increase.

The proposed algorithm in §2 uses a 27 dimensional vector in its calculation of the zeta score at each voxel location. The vector can be of any size, and although used to help incorporate local anatomical information in the decision whether or not to label a voxel as abnormal in this case, additional sequences can be used making the algorithm multispectral. All that is required is for the different imaging modalities to be in spatial register. Following on from this, the algorithm can be applied to other lesions that are not vascular. What is required is a sequence that differentiates between healthy and injured tissue and an appropriate normal reference data set. For example, it may be possible to translate the algorithm to identify demyelinating plaques, such as in multiple sclerosis, by using a T2 FLAIR sequence (FLuid Attenuation Inverse Recovery).

In §3, different algorithms were assessed on the amount of perturbation the presence of a lesion produced on the spatial normalisation result. Ideally the spatial normalisation parameters on two identical brains except for the presence and absence of a lesion should be identical. Any error that is incurred is secondary to how the algorithm handles the presence of a lesion, with the result deteriorating the subsequent statistical inference either through an increase in noise, or worse, bias. Therefore it is beneficial to identify an algorithm with as little error as possible. The simulations in §3 have shown that the unified segmentation-normalisation with light regularisation and enantiomorphic correction (ULE) to have the smallest amount of error measured as the root mean squared difference for small to medium sized lesions. For lesions greater than 24000mm^3 there was no significant difference compared with the unified segmentation-normalisation with light regularisation and cost function masking (ULC). From the point of view of automation, a blanket use of ULE would be preferable, since any brain image with a lesion

greater than 24000mm^3 would not incur any less error if ULC had been employed. Although it is possible to use the unsupervised lesion segmentation routine described in §2 to assess the lesion volume first to stratify the process, such a step would appear excessive in this circumstance.

The techniques that necessitate lesion segmentation prior to spatial normalisation may benefit from an unsupervised lesion segmentation routine. The technique described in §2 can be used in combination with ULE, to improve both lesion segmentation and spatial normalisation by utilising an iterative process, where all spatial normalisation except for the very first instance is via ULE. Since the computation time of lesion segmentation and spatial normalisation is within the order of 8 minutes and 10 minutes respectively, a series of three iterations will still be on average faster than what is typically expected from manual segmentation alone (Andersen, Rapcsak, and Beeson 2010).

7.3 Inference and prediction

Chapters §4 and §5 examined the concerns regarding mass univariate techniques in brain mapping, where the spatial dependencies within the data were not explicitly modelled. The simulations suggested that the biases present within the dataset resulted in mislocalisation of the critical region for both single and multiple loci models. This of course raises concerns regards past lesion studies that have employed the mass univariate technique and whether a multivariate approach described in §5 would derive the same localisation after explicitly modelling for spatial dependencies. It would be preferable to perform a simulation on a clinical scenario and compare the results of uni and multivariate analysis, however the structure-functional relationship is not known and therefore it is impossible to ascertain which approach is correct.

We are thus left with the prediction accuracy of the final model and supportive evidence from alternative brain mapping techniques as our guide.

Chapter §6 applied the multivariate technique of Support Vector Machines (SVM) to the classification task of differentiating those patients who had experienced an acute vascular injury, as either dependent or independent (determined by their Barthel Index) solely from their admission diffusion weighted MRI brain scan. The SVM model was trained to predict the Barthel Index (BI) a patient would achieve within 3 weeks of admission. Despite the small sample size of 93 patients and a dimensionality mismatch of 93 to 25122 the results were promising with prediction accuracies between 64% to 75%. Although it has been shown that the BI within 3 weeks provides some prognostic information regards a patient's level of dependence at 6 months, it is unlikely for this value to be particularly stable at this stage as many post stroke patients will receive neuro-rehabilitation for a number of weeks after injury. Therefore the accuracy of the model at predicting a patient's BI within 3 weeks may be part secondary to the variability of the outcome measure itself.

When the weights of the SVM model are assessed the voxels that are attributed the greatest importance for predicting dependence cluster around the corona radiata, a region where both primary and secondary motor efferents pass through. Although other regions of the brain were also identified by the negative weights, this finding would be consistent with the BI having a significant motor emphasis in its method of assessing a patient's level of functioning.

The Barthel Index, however, is a summary measure of a patient's level of functionality which will incorporate both motor and cognitive skills, and it is not possible to discern whether the final score is secondary purely to a motor deficit or from a multifactorial cause. This lack of specificity in the outcome measure may impinge on the attainable accuracy by the multivariate

model. Consequently by modelling more restricted functions such as upper limb movement or language impairment such as those found in the National Institute of Health Stroke Score (NIHSS) a better performing model may be obtained. Additionally increasing the specificity of the outcome measure may be considered clinically more useful as it may provide better insight into the specific needs of each patient.

7.4 Future work

In §5 the benefits of multivariate modelling over mass univariate techniques with respects to exploring the functional architecture of the brain were examined. While in §6 the predictive qualities of the multivariate model on a real clinical problem were assessed. It is clear that these multivariate techniques have the potential to improve our understanding of the brain and benefit clinical practice, but currently remain in their infancy. The methodological techniques discussed in chapters §2 to §4 describe a potential path along which these models can be further developed. With the development of Hyper Acute Stroke Units (HASU) in the United Kingdom, and regional rehabilitation centres, paired datasets of imaging and outcome measures will become more readily available. By automating the image pre-processing stage the path to conducting studies with sample sizes in the excess of hundreds or possibly thousands of patients is hopefully simplified. The implementation of these techniques are relatively simple, as they do not require any specialised equipment to run them. Although Support Vector Machines have been used as our example technique in §6, the simulations from §5 would suggest a Bayesian approach such as with Flexible Bayesian Modelling (FBM) is superior. Moreover, a Bayesian model would allow a clearer quantification of the level of confidence in the prediction obtain for each patient which is not possible with a SVM model.

With larger datasets future work would aim at re-examining past lesion studies through the lens of multivariate models, looking for possible inconsistencies with past studies that have utilised a univariate technique and alternative brain mapping techniques. As a clinical tool, further work using larger sample sizes as well as outcomes with greater specificity is necessary to improve the predictive accuracy of these models.

8 Appendix

8.1 Appendix A

Author	Sensitivity (%)			Specificity (%)			Similarity Index		
	Max.	Min.	Central tendency	Max.	Min.	Central tendency	Max.	Min.	Central tendency
Mah	100.00	80.29	96.02	99.98	98.02	99.79	0.90	0.41	0.73
Bhanu Prakash	99.92	13.48	72.70	100.00	97.48	99.66	0.96	0.22	0.67
Gupta	-	-	79.30	-	-	99.30	-	-	0.68
Hevai-Montiel	-	-	-	-	-	-	0.83	0.29	0.54

Figure 8.1 - Descriptive statistics of stroke lesion segmentation results on diffusion weighted MR images.

The measure of central tendency was the mean for all except Mah, which was the median.

	%											
BA	5	10	15	20	25	30	35	40	45	50	55	60
1	26.68	26.61	20.78	14.70	32.06	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	24.25	18.87	18.97	23.49	14.42	14.00	18.44	7.21	8.94	4.47	7.48	NaN
3	22.45	21.63	20.10	14.70	16.61	19.80	11.66	11.66	10.20	5.66	8.49	10.95
4	28.35	30.13	22.45	16.61	18.87	27.86	NaN	8.25	3.46	NaN	NaN	NaN
5	13.42	14.70	NaN	108.92	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6	28.07	28.71	33.35	13.56	14.70	11.66	NaN	NaN	NaN	NaN	NaN	NaN
7	24.74	14.97	14.00	7.48	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	12.33	12.33	7.48	NaN	NaN	11.66	NaN	NaN	NaN	NaN	NaN	NaN
9	20.69	14.14	7.48	8.49	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	16.61	9.17	8.49	8.72	8.72	6.63	NaN	NaN	NaN	NaN	NaN	NaN
11	15.62	14.00	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	10.00	6.63	10.20	5.66	16.49	12.96	15.62	10.20	8.94	NaN	10.39	NaN
18	8.94	10.95	6.32	8.00	18.55	10.95	NaN	NaN	NaN	NaN	NaN	NaN
19	7.21	6.93	9.38	8.72	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	21.26	19.08	14.97	14.70	26.31	NaN	NaN	NaN	NaN	NaN	NaN	NaN
21	14.56	16.12	14.42	13.11	14.42	12.81	14.14	12.00	10.77	24.17	2.83	16.61
22	17.66	17.32	12.81	13.27	16.12	15.75	14.56	18.22	14.42	12.33	9.17	6.00
23	25.46	23.41	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
24	14.97	13.56	14.70	10.77	11.66	4.47	6.63	8.25	7.21	3.46	NaN	8.72
25	12.81	7.21	4.00	4.00	4.90	NaN	10.20	NaN	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	8.72	12.33	12.33	7.48	6.00	7.48	NaN	NaN	NaN	2.00	NaN	NaN
28	18.33	8.49	8.25	4.47	2.83	NaN	NaN	NaN	NaN	NaN	NaN	NaN
29	3.46	46.73	3.46	4.90	3.46	NaN	NaN	NaN	NaN	NaN	NaN	NaN
30	17.89	8.49	10.95	6.32	7.48	5.66	2.00	12.81	NaN	NaN	NaN	NaN
32	19.08	8.00	10.39	9.17	8.25	17.32	4.47	8.25	6.63	8.94	NaN	NaN
34	17.20	14.56	19.60	15.62	11.66	14.56	6.93	4.47	2.83	4.47	6.32	6.93
35	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
36	16.12	14.70	11.83	NaN	7.48	NaN	NaN	NaN	NaN	NaN	NaN	NaN
37	23.75	24.66	20.10	18.55	12.33	NaN	NaN	NaN	NaN	NaN	NaN	NaN
38	30.27	24.49	22.89	24.41	24.17	35.44	24.08	15.23	16.97	17.20	8.25	17.89
39	25.38	26.38	25.38	25.38	21.91	25.77	22.72	17.44	20.10	19.80	8.72	6.00
40	28.50	18.97	16.25	16.61	20.40	13.42	12.00	6.63	4.90	8.49	3.46	4.47
41	11.49	12.33	11.83	16.97	18.55	15.75	9.17	11.83	6.32	14.00	10.20	11.83
42	20.10	10.39	12.96	12.96	17.32	14.70	11.83	12.33	8.25	8.25	8.94	11.49
43	12.17	15.23	14.00	14.00	12.81	20.88	22.09	9.17	15.75	14.70	15.62	8.25
44	18.55	19.90	15.36	18.97	17.32	15.23	15.62	12.17	10.77	14.70	8.25	12.17
45	22.72	26.08	18.00	15.10	16.61	22.09	10.95	16.12	24.00	4.00	7.21	22.89
46	25.61	19.39	14.70	7.21	7.21	29.60	NaN	NaN	NaN	NaN	NaN	NaN
47	33.65	34.23	38.11	36.82	27.20	29.60	25.46	25.38	7.21	14.14	18.97	9.38
48	6.32	4.47	5.66	7.21	6.00	5.66	4.00	7.48	6.93	11.83	7.21	10.77

Figure 8.2 - Table of mislocalisation for each Brodmann area calculated from the dataset of 581 lesion masks.

The value at each Brodmann area was calculated by labelling the stack of 581 lesioned volumes as being “affected” or “unaffected” depending on whether or not a minimum percentage volume of the Brodmann area fell within the lesion in each volume, running a standard voxel-wise Fisher’s exact test-based mass-univariate analysis on the two groups, and identifying the centre of the resultant significant cluster, identified by the asymptotic p value threshold at a Bonferroni corrected $p < 0.01$.

8.2 Appendix B

8.2.1 Statistical Parametric Mapping 5 (SPM5) settings

The SPM5 software suite consists of a collection of modifiable functions whose settings are specified in a spm_defaults file.

8.2.1.1 SPM5 co-registration defaults

```
coreg.estimate.cost_fun = 'nmi';
coreg.estimate.sep      = [4 2];
coreg.estimate.tol      = [0.02 0.02 0.02 0.001 0.001
0.001 0.01 0.01 0.01 0.001 0.001 0.001];
coreg.estimate.fwhm     = [7 7];
coreg.write.interp      = 1;
coreg.write.wrap        = [0 0 0];
coreg.write.mask        = 0;
```

8.2.1.2 SPM5 preproc (unified segmentation-normalisation routine)

```
preproc.tpm     = char(...
    fullfile(spm('Dir'),'tpm','grey.nii'),...
    fullfile(spm('Dir'),'tpm','white.nii'),...
    fullfile(spm('Dir'),'tpm','csf.nii')); % Prior
probability maps
preproc.ngaus    = [2 2 2 4];     % Gaussians per class
preproc.warpreg  = 1;             % Warping Regularisation
preproc.warpco   = 25;            % Warp Frequency Cutoff
preproc.biasreg  = 0.0001;        % Bias regularisation
preproc.biasfwhm = 75;            % Bias FWHM
preproc.regtype  = 'mni';         % Affine Regularisation
preproc.samp     = 3;             % Sampling distance
```

8.2.2 Statisitical Parametric Mapping 8 (SPM8) settings

The SPM8 software suite consists of a collection of modifiable functions that can be scripted in the form of “jobs”. The location of the function / tool within the suite is specified prior to the setting details.

8.2.2.1 SPM8 Segment

```
[Tools -> Spatial -> Segment]
%------------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%------------------------------------------------------------
matlabbatch{1}.spm.spatial.preproc.data = '<Source_
Image>';
matlabbatch{1}.spm.spatial.preproc.output.GM = [0 0 1];
matlabbatch{1}.spm.spatial.preproc.output.WM = [0 0 1];
matlabbatch{1}.spm.spatial.preproc.output.CSF = [0 0 0];
matlabbatch{1}.spm.spatial.preproc.output.biascor = 1;
matlabbatch{1}.spm.spatial.preproc.output.cleanup = 0;
matlabbatch{1}.spm.spatial.preproc.opts.tpm = {
                                    '\spm8\tpm\grey.nii'
                                    '\spm8\tpm\white.nii'
                                    '\spm8\tpm\csf.nii'
                                                    };
matlabbatch{1}.spm.spatial.preproc.opts.ngaus = [2
                                                  2
                                                  2
                                                  4];
matlabbatch{1}.spm.spatial.preproc.opts.regtype = 'mni';
matlabbatch{1}.spm.spatial.preproc.opts.warpreg = 1;
matlabbatch{1}.spm.spatial.preproc.opts.warpco = 25;
matlabbatch{1}.spm.spatial.preproc.opts.biasreg = 0.0001;
matlabbatch{1}.spm.spatial.preproc.opts.biasfwhm = 60;
matlabbatch{1}.spm.spatial.preproc.opts.samp = 3;
matlabbatch{1}.spm.spatial.preproc.opts.msk = {''};
```

8.2.2.2 SPM8 Co-register

```
[Tools -> Spatial -> Coregistration : Estimate and Write]
```

```
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.spatial.coreg.estwrite.ref =
'<Reference_Image>';
matlabbatch{1}.spm.spatial.coreg.estwrite.source =
'<Source_Image>';
matlabbatch{1}.spm.spatial.coreg.estwrite.other = {''};
matlabbatch{1}.spm.spatial.coreg.estwrite.eoptions.cost_
fun = 'nmi';
matlabbatch{1}.spm.spatial.coreg.estwrite.eoptions.sep =
[4 2];
matlabbatch{1}.spm.spatial.coreg.estwrite.eoptions.tol =
[0.02 0.02 0.02 0.001 0.001 0.001 0.01 0.01 0.01 0.001
0.001 0.001];
matlabbatch{1}.spm.spatial.coreg.estwrite.eoptions.fwhm =
[7 7];
matlabbatch{1}.spm.spatial.coreg.estwrite.roptions.interp
= 1;
matlabbatch{1}.spm.spatial.coreg.estwrite.roptions.wrap =
[0 0 0];
matlabbatch{1}.spm.spatial.coreg.estwrite.roptions.mask =
0;
matlabbatch{1}.spm.spatial.coreg.estwrite.roptions.prefix
= 'r';
```

8.2.2.3 SPM8 Normalise

```
[Tools -> Spatial -> Normalise -> Estimate and Write]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.spatial.normalise.estwrite.subj.source
= '<Source_Image>';
matlabbatch{1}.spm.spatial.normalise.estwrite.subj.wtsrc =
'';
matlabbatch{1}.spm.spatial.normalise.estwrite.subj.
resample = '<UNDEFINED>';
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
```

```
template = '<UNDEFINED>';
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
weight = '';
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
smosrc = 8;
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
smoref = 0;
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
regtype = 'mni';
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
cutoff = 25;
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.
nits = 16;
matlabbatch{1}.spm.spatial.normalise.estwrite.eoptions.reg
= 1;
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.
preserve = 0;
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.bb
= [-78 -112 -50

78 76 85];
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.vox
= [2 2 2];
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.
interp = 1;
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.
wrap = [0 0 0];
matlabbatch{1}.spm.spatial.normalise.estwrite.roptions.
prefix = 'w';
```

8.2.2.4 SPM8 Reorient

```
if length(vx)<3
    vx=[vx vx vx];
end

% If no arguments, then prompt for images
%PP = spm_get([1 Inf],'*.img','Select files to reorient');
```

```
% Get information about the image volumes
VV = spm_vol(filename);

for V=VV', % Loop over images
    % The corners of the current volume
    d = V.dim(1:3);
    c = [    1    1    1    1
        1    1    d(3) 1
        1    d(2) 1    1
        1    d(2) d(3) 1
        d(1) 1    1    1
        d(1) 1    d(3) 1
        d(1) d(2) 1    1
        d(1) d(2) d(3) 1]';

    % The corners of the volume in mm space
    tc = V.mat(1:3,1:4)*c;
    if spm_flip_analyze_images, tc(1,:) = -tc(1,:); end;

    % Max and min co-ordinates for determining a bounding-
box
    mx = round(max(tc,[],2)');
    mn = round(min(tc,[],2)');

    % Translate so that minimum moves to [1,1,1]
    % This is the key bit for changing voxel sizes,
    % output orientations etc.
    mat = spm_matrix(mn)*diag([vx 1])*spm_matrix(-[1 1
1]);

    % Dimensions in mm
    dim = ceil((mat\[mx 1]')');

    % Output image based on information from the original
    VO               = V;

    % Create a filename for the output image (prefixed by
'r')
    [lpath,name,ext] = fileparts(V.fname);
```

```
    VO.fname         = fullfile(lpath,['ro_' name ext]);

    % Dimensions of output image
    VO.dim(1:3)      = dim(1:3);

    % Voxel-to-world transform of output image
    if spm_flip_analyze_images, mat = diag([-1 1 1
1])*mat; end;
    VO.mat           = mat;

    % Initialise plot of how far reslicing has gone
    %spm_progress_bar('Init',dim(3),'reslicing...','planes
completed');

    % Create .hdr and open output .img
    VO = spm_create_vol(VO);

    for i=1:dim(3), % Loop over slices of output image

        % Mapping from slice i of the output image,
        % to voxels of the input image
        M   = inv(spm_matrix([0 0 -i])*inv(VO.mat)*V.mat);

        % Extract this slice according to the mapping
        img = spm_slice_vol(V,M,dim(1:2),-6);

        % Write this slice to output image
        spm_write_plane(VO,img,i);

        % Update the progress bar
        %spm_progress_bar('Set',i);

    end; % End loop over output slices

    % Get rid of the progress bar
    %spm_progress_bar('Clear');
end; % End loop over images
return; % Done
```

8.2.3 SPM8 settings used in §2

The assessment of the enantiomorphic normalisation modification against other existing techniques required the default SPM8 function settings to be modified depending on the method.

The various scripts were performed using the default spm settings for FMRI [spm('defaults', 'FMRI')] and passed through a batching pipeline by calling the native "jobman" function within SPM8.

8.2.3.1 Unified Segment [ULP, ULPD, ULC, ULCD, ULE, ULED]

```
[Tools -> Spatial -> Segment]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.spatial.preproc.data = '<source_
image>';
matlabbatch{1}.spm.spatial.preproc.output.GM = [0 0 1];
matlabbatch{1}.spm.spatial.preproc.output.WM = [0 0 1];
matlabbatch{1}.spm.spatial.preproc.output.CSF = [0 0 0];
matlabbatch{1}.spm.spatial.preproc.output.biascor = 0;
matlabbatch{1}.spm.spatial.preproc.output.cleanup = 0;
matlabbatch{1}.spm.spatial.preproc.opts.tpm = {
                                      '\spm8\tpm\grey.nii'
                                      '\spm8\tpm\white.nii'
                                      '\spm8\tpm\csf.nii'
                                                 };
matlabbatch{1}.spm.spatial.preproc.opts.ngaus = [2
                                                  2
                                                  2
                                                  4];
matlabbatch{1}.spm.spatial.preproc.opts.regtype = 'mni';
matlabbatch{1}.spm.spatial.preproc.opts.warpreg = 1;
matlabbatch{1}.spm.spatial.preproc.opts.warpco = 25;
matlabbatch{1}.spm.spatial.preproc.opts.biasreg =
```

```
'<regularisation_setting>';
matlabbatch{1}.spm.spatial.preproc.opts.biasfwhm = 60;
matlabbatch{1}.spm.spatial.preproc.opts.samp = 3;
if isempty(mname);
    matlabbatch{1}.spm.spatial.preproc.opts.msk = {''};
else
    matlabbatch{1}.spm.spatial.preproc.opts.msk =
'<mask>';
end
```

8.2.3.2 Create deformation field from sn file [ULP, ULPD, ULC, ULCD, ULE, ULED]

```
[Util -> Deformations -> Import _sn.mat]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.util.defs.comp{1}.sn2def.matname =
'<sn_parameter_file>';
matlabbatch{1}.spm.util.defs.comp{1}.sn2def.vox = [NaN NaN
NaN];
matlabbatch{1}.spm.util.defs.comp{1}.sn2def.bb = [NaN NaN
NaN
                                                  NaN NaN
NaN];
matlabbatch{1}.spm.util.defs.ofname = '<output_name>';
matlabbatch{1}.spm.util.defs.fnames = '';
matlabbatch{1}.spm.util.defs.savedir.saveusr = '<output_
path>';
matlabbatch{1}.spm.util.defs.interp = '<interpolation_
setting>'
```

8.2.3.3 Import tissue classes for use with DARTEL [ULPD, ULCP, ULED]

```
[Tools -> DARTEL Tools -> Initial Import]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.tools.dartel.initial.matnames = '<sn_
```

```
parameter_file>';
matlabbatch{1}.spm.tools.dartel.initial.odir = '<output_
path>';
matlabbatch{1}.spm.tools.dartel.initial.bb = [NaN NaN NaN
                                               NaN NaN
NaN];
matlabbatch{1}.spm.tools.dartel.initial.vox = 1.5;
matlabbatch{1}.spm.tools.dartel.initial.image = 0;
matlabbatch{1}.spm.tools.dartel.initial.GM = 1;
matlabbatch{1}.spm.tools.dartel.initial.WM = 1;
matlabbatch{1}.spm.tools.dartel.initial.CSF = 0;
```

8.2.3.4 Create DARTEL Templates [ULPD, ULCD, ULED]

```
[Tools -> DARTEL Tools -> Run DARTEL (create Templates)]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.tools.dartel.warp.images = {
                                                {
                                                [rc1 ',1']
                                                [rc2 ',1']
                                                }
                                                }';
matlabbatch{1}.spm.tools.dartel.warp.settings.template =
'Template';
matlabbatch{1}.spm.tools.dartel.warp.settings.rform = 0;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(1).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(1).
rparam = [4 2 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(1).K =
0;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(1).
slam = 16;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(2).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(2).
rparam = [2 1 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(2).K =
0;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(2).
slam = 8;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(3).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(3).
rparam = [1 0.5 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(3).K =
1;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(3).
slam = 4;
```

```
matlabbatch{1}.spm.tools.dartel.warp.settings.param(4).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(4).
rparam = [0.5 0.25 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(4).K =
2;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(4).
slam = 2;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(5).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(5).
rparam = [0.25 0.125 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(5).K =
4;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(5).
slam = 1;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(6).its
= 3;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(6).
rparam = [0.25 0.125 1e-006];
matlabbatch{1}.spm.tools.dartel.warp.settings.param(6).K =
6;
matlabbatch{1}.spm.tools.dartel.warp.settings.param(6).
slam = 0.5;
matlabbatch{1}.spm.tools.dartel.warp.settings.optim.lmreg
= 0.01;
matlabbatch{1}.spm.tools.dartel.warp.settings.optim.cyc =
3;
matlabbatch{1}.spm.tools.dartel.warp.settings.optim.its =
3;
```

8.2.3.5 Transform the DARTEL flow field into MNI space

```
[Tools -> DARTEL Tools -> Normalise to MNI space]

% template      : the last template created by dartel to
                  normalise into mni
% tmptype       : the method used to create the tissue
                  class
```

```
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.spatial.normalise.est.subj.source =
{[template ',1']};
matlabbatch{1}.spm.spatial.normalise.est.subj.wtsrc = '';
if strcmp('UDartel',tmptype)
    matlabbatch{1}.spm.spatial.normalise.est.eoptions.
template = {'/spm8/aprior/grey.nii,1'};
elseif strcmp('NDartel', tmptype)
    matlabbatch{1}.spm.spatial.normalise.est.eoptions.
template = {'/spm8/tmp/grey.nii,1'};
end
matlabbatch{1}.spm.spatial.normalise.est.eoptions.weight =
'';
matlabbatch{1}.spm.spatial.normalise.est.eoptions.smosrc =
8;
matlabbatch{1}.spm.spatial.normalise.est.eoptions.smoref =
0;
matlabbatch{1}.spm.spatial.normalise.est.eoptions.regtype
= 'mni';
matlabbatch{1}.spm.spatial.normalise.est.eoptions.cutoff =
25;
matlabbatch{1}.spm.spatial.normalise.est.eoptions.nits =
16;
matlabbatch{1}.spm.spatial.normalise.est.eoptions.reg = 1;
```

8.2.3.6 Create deformation field from DARTEL flow field [ULPD, ULCD, ULED]

```
[Util -> Deformations -> DARTEL flow]

% dart2mni_sn   : affine transform from dartel space to
                  mni
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.util.defs.comp{1}.dartel.flowfield =
'<flowfield>';
matlabbatch{1}.spm.util.defs.comp{1}.dartel.times = [1 0];
```

```
matlabbatch{1}.spm.util.defs.comp{1}.dartel.K = 6;
matlabbatch{1}.spm.util.defs.comp{2}.sn2def.matname =
{[dart2mni_sn]};
matlabbatch{1}.spm.util.defs.comp{2}.sn2def.vox = [NaN NaN
NaN];
matlabbatch{1}.spm.util.defs.comp{2}.sn2def.bb = [NaN NaN
NaN
                                                   NaN NaN
NaN];
matlabbatch{1}.spm.util.defs.ofname = oname;
matlabbatch{1}.spm.util.defs.fnames = '';
matlabbatch{1}.spm.util.defs.savedir.saveusr = '<output_
path>';
matlabbatch{1}.spm.util.defs.interp = <interpolation>;
```

8.2.3.7 New Segment [NSPD, NWPD, NSED, NWED]

```
[Tools -> New Segment]
%-----------------------------------------------------------
% Job configuration created by cfg_util (rev $Rev: 3944 $)
%-----------------------------------------------------------
matlabbatch{1}.spm.tools.preproc8.channel.vols = {<source_
image>};
matlabbatch{1}.spm.tools.preproc8.channel.biasreg = regu;
% default==0.0001
matlabbatch{1}.spm.tools.preproc8.channel.biasfwhm = 60;
matlabbatch{1}.spm.tools.preproc8.channel.write = [0 0];
matlabbatch{1}.spm.tools.preproc8.tissue(1).tpm = {'\spm8\
toolbox\Seg\TPM.nii,1'};
matlabbatch{1}.spm.tools.preproc8.tissue(1).ngaus = 2;
matlabbatch{1}.spm.tools.preproc8.tissue(1).native = [0
1];
matlabbatch{1}.spm.tools.preproc8.tissue(1).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(2).tpm = {'\spm8\
toolbox\Seg\TPM.nii,2'};
matlabbatch{1}.spm.tools.preproc8.tissue(2).ngaus = 2;
matlabbatch{1}.spm.tools.preproc8.tissue(2).native = [0
1];
```

```
matlabbatch{1}.spm.tools.preproc8.tissue(2).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(3).tpm = {'\spm8\
toolbox\Seg\TPM.nii,3'};
matlabbatch{1}.spm.tools.preproc8.tissue(3).ngaus = 2;
matlabbatch{1}.spm.tools.preproc8.tissue(3).native = [0
1];
matlabbatch{1}.spm.tools.preproc8.tissue(3).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(4).tpm = {'\spm8\
toolbox\Seg\TPM.nii,4'};
matlabbatch{1}.spm.tools.preproc8.tissue(4).ngaus = 3;
matlabbatch{1}.spm.tools.preproc8.tissue(4).native = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(4).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(5).tpm = {'\spm8\
toolbox\Seg\TPM.nii,5'};
matlabbatch{1}.spm.tools.preproc8.tissue(5).ngaus = 4;
matlabbatch{1}.spm.tools.preproc8.tissue(5).native = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(5).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(6).tpm = {'\spm8\
toolbox\Seg\TPM.nii,6'};
matlabbatch{1}.spm.tools.preproc8.tissue(6).ngaus = 2;
matlabbatch{1}.spm.tools.preproc8.tissue(6).native = [0
0];
matlabbatch{1}.spm.tools.preproc8.tissue(6).warped = [0
0];
matlabbatch{1}.spm.tools.preproc8.warp.reg = 4;
matlabbatch{1}.spm.tools.preproc8.warp.affreg = 'mni';
matlabbatch{1}.spm.tools.preproc8.warp.samp = 3;
matlabbatch{1}.spm.tools.preproc8.warp.write = [1 1];
```

9 Bibliography

Adèr H. Advising on research methods : a consultant's companion. Huizen Netherlands: Johannes van Kessel Pub.; 2008.

Adolphs R, Tranel D, Damasio H, Damasio AR. Fear and the human amygdala. J. Neurosci. 1995; 15: 5879–5891.

Anbeek P, Vincken KL, van Bochove GS, van Osch MJP, van der Grond J. Probabilistic segmentation of brain tissue in MR imaging. Neuroimage 2005; 27: 795–804.

Anbeek P, Vincken KL, van Osch MJP, Bisschops RHC, van der Grond J. Automatic segmentation of different-sized white matter lesions by voxel probability estimation. Med.Image Anal. 2004; 8: 205–215.

Andersen SM, Rapcsak SZ, Beeson PM. Cost function masking during normalization of brains with focal lesions: Still a necessity? NeuroImage 2010; 53: 78–84.

Ashburner J, Friston K. Multimodal Image Coregistration and Partitioning - A Unified Framework. Neuroimage 1997; 6: 209–217.

Ashburner J, Friston KJ. Nonlinear Spatial Normalization Using Basis Functions. Hum.Brain Mapp. 1999; 7: 254–266.

Ashburner J, Friston KJ. Voxel-based morphometry - the methods. Neuroimage 2000; 11: 805–821.

Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005; 26: 839–851.

Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage 2007; 38: 95–113.

Ashton EA, Takahashi C, Berg MJ, Goodman A, Totterman S, Ekholm S. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. Journal of Magnetic Resonance Imaging 2003; 17: 300–308.

Atlas SA. The renin-angiotensin aldosterone system: pathophysiological role and pharmacologic inhibition. J Manag Care Pharm 2007; 13: 9–20.

Aue T, Lavelle LA, Cacioppo JT. Great expectations: what can fMRI research tell us about psychological phenomena? Int J Psychophysiol 2009; 73: 10–16.

Baird AE, Benfield A, Schlaug G, Siewert B, Lövblad KO, Edelman RR, et al. Enlargement of human cerebral ischemic lesion volumes measured by diffusion-weighted magnetic resonance imaging. Ann. Neurol. 1997; 41: 581–589.

Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet 1991; 337: 1521–1526.

Bang OY, Lee PH, Heo KG, Joo US, Yoon SR, Kim SY. Specific DWI lesion patterns predict prognosis after acute ischaemic stroke within the MCA territory. Journal of Neurology, Neurosurgery & Psychiatry 2005; 76: 1222–1228.

Barbas H, Mesulam MM. Organization of afferent input to subdivisions of area 8 in the rhesus monkey. J. Comp. Neurol. 1981; 200: 407–431.

Barber PA, Darby DG, Desmond PM, Yang Q, Gerraty RP, Jolley D, et al. Prediction of stroke outcome with echoplanar perfusion- and diffusion-weighted MRI. Neurology 1998; 51: 418–426.

Barrett WA, Mortensen EN. Interactive live-wire boundary extraction. Medical image analysis 1997; 1: 331–341.

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, et al. Voxel-based lesion-symptom mapping. Nat.Neurosci. 2003; 6: 448–449.

Bates JF, Goldman-Rakic PS. Prefrontal connections of medial motor areas in the rhesus monkey. J. Comp. Neurol. 1993; 336: 211–228.

Bergeest JP, Jager F. A Comparison of Five Methods for Signal Intensity Standardization in MRI. In: Bildverarbeitung für die Medizin 2008. Algorithmen — Systeme — Anwendungen Proceedings des Workshops vom 6. bis 8. April 2008 in Berlin. Springer Berlin Heidelberg; 2008.

Berman MG, Jonides J, Nee DE. Studying mind and brain with fMRI. Social Cognitive and Affective Neuroscience 2006; 1: 158–161.

Bhanu Prakash KN, Gupta V, Jianbo H, Nowinski WL. Automatic processing of diffusion-weighted ischemic stroke images based on divergence measures: slice and hemisphere identification, and stroke region segmentation. International Journal of Computer Assisted Radiology and Surgery 2008; 3: 559–570.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. J. Neurosci. 1997; 17: 353–362.

Bird CM, Malhotra P, Parton A, Coulthard E, Rushworth MFS, Husain M. Visual neglect after right posterior cerebral artery infarction. Journal of Neurology, Neurosurgery & Psychiatry 2006; 77: 1008–1012.

Boser BE, Guyon I., Vapnik V. A Training Algorithm for Optimal Margin Classifiers. Pittsburgh, Pennsylvania, USA: ACM; 1992. p. 144–152.

Bradley PS, Fayyad UM. Refining initial points for k-means clustering. In: Proceedings of the 15th International Conference on Machine Learning. San Francisco, USA: 1998. p. 91–99.

Breiman L, Spector P. Submodel selection and evaluation in regression. The X-random case. International Statistical Review 1992; 60: 291–319.

Breiman L. Heuristics of instability and stabilization in model selection. The Annals of Statistics 1996; 24: 2350–2383.

Brett M, Leff AP, Rorden C, Ashburner J. Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. Neuroimage 2001; 14: 486–500.

Broca P. REMARQUES SUR LE SIÉGE DE LA FACULTÉ DU LANGAGE ARTICULÉ, SUIVIES D'UNE OBSERVATION D'APHÉMIE (PERTE DE LA PAROLE). Bulletin de la Société Anatomique 1861; 6: 330–357.

Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 2009; 10: 186–198.

Calder AJ, Keane J, Manes F, Antoun N, Young AW. Impaired recognition and experience of disgust following brain injury. Nature Neuroscience 2000; 3: 1077–1078.

Callaway EM. Local circuits in primary visual cortex of the macaque monkey. Annu. Rev. Neurosci. 1998; 21: 47–74.

Catani M, Dell'acqua F, Vergani F, Malik F, Hodge H, Roy P, et al. Short frontal lobe connections of the human brain. Cortex 2012; 48: 273–291.

Catani M, Howard RJ, Pajevic S, Jones DK. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. Neuroimage 2002; 17: 77–94.

Catani M, Thiebaut de Schotten M. A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 2008; 44: 1105–1132.

Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. 2001

Chen CL, Tang FT, Chen HC, Chung CY, Wong MK. Brain lesion size and location: effects on motor recovery and functional outcome in stroke patients. Arch Phys Med Rehabil 2000; 81: 447–452.

Chen H, Tino P, Yao X. Probabilistic classification vector machines. IEEE Trans Neural Netw 2009; 20: 901–914.

Cho S-H, Kim DG, Kim D-S, Kim Y-H, Lee C-H, Jang SH. Motor outcome according to the integrity of the corticospinal tract determined by diffusion tensor tractography in the early stage of corona radiata infarct. Neurosci. Lett. 2007; 426: 123–127.

Chodorowski A, Mattsson U, Langille M, Hamarneh G, Fitzpatrick JM, Reinhardt JM. Color lesion boundary detection using live wire. San Diego, CA, USA: 2005. p. 1589–1596.

Colliot O, Bernasconi N, Khalili N, Antel SB, Naessens V, Bernasconi A. Individual voxel-based analysis of gray matter in focal cortical dysplasia. Neuroimage 2006; 29: 162–171.

Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002; 24: 603–619.

Cortes C, Vapnik V. Support-vector networks. Machine Learning 1995; 20: 273–297.

Cover T, Hart P. Nearest neighbor pattern classification. Information Theory, IEEE Transactions on 1967; 13: 21–27.

Crinion J, Ashburner J, Leff A, Brett M, Price C, Friston K. Spatial normalization of lesioned brains: Performance evaluation and impact on fMRI analyses. Neuroimage 2007; 37: 866–875.

Cristianini N, Campbell C, Shawe-Taylor J. Dynamically Adapting Kernels in Support Vector Machines. Advances in neural information processing systems 1998: 204–210.

Damasio AR, Damasio H, Chui HC. Neglect following damage to frontal lobe or basal ganglia. Neuropsychologia 1980; 18: 123–132.

Damasio H, Damasio AR. Lesion analysis in neuropsychology. New York: Oxford University Press; 1989.

Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology 1945; 26: 297–302.

Dougherty RF, Ben-Shachar M, Bammer R, Brewer AA, Wandell BA. Functional organization of human occipital-callosal fiber tracts. Proc. Natl. Acad. Sci. U.S.A. 2005; 102: 7350–7355.

Downing PE. A Cortical Area Selective for Visual Processing of the Human Body. Science 2001; 293: 2470–2473.

Dronkers NF, Colliot O, Iba-Zizen MT, Cabanis EA. Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. Brain 2007; 130: 1432–1441.

Dronkers NF, Wilkins DP, Van Valin RD, Redfern BB, Jaeger JJ. Lesion analysis of the brain areas involved in language comprehension. Cognition 2004; 92: 145–177.

Duda RO, Hart PE. Pattern classification and scene analysis. New York: John Wiley and Sons; 1973.

Duncan PW, Samsa GP, Weinberger M, Goldstein LB, Bonito A, Witter DM, et al. Health status of individuals with mild stroke. Stroke 1997; 28: 740–745.

Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1994.

Efron B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. JASA 1983; 78: 316–331.

Egly R, Driver J, Rafal RD. Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. Journal of Experimental Psychology: General 1994; 123: 161–177.

Epstein CM, Schwartzberg DG, Davey KR, Sudderth DB. Localizing the site of magnetic brain stimulation in humans. Neurology 1990; 40: 666–670.

Van Everdingen KJ, van der Grond J, Kappelle LJ, Ramos LM, Mali WP. Diffusion-weighted magnetic resonance imaging in acute stroke. Stroke 1998; 29: 1783–1790.

Fair DA, Cohen AL, Power JD, Dosenbach NUF, Church JA, Miezin FM, et al. Functional Brain Networks Develop from a 'Local to Distributed' Organization. PLoS Computational Biology 2009; 5: e1000381.

Falcao AX, Udupa JK, Miyazawa FK. An ultra-fast user-steered image segmentation paradigm: live wire on the fly. Medical Imaging, IEEE Transactions on 2000; 19: 55–62.

Falcao AX, Udupa JK, Samarasekera S, Sharma S, Hirsch BE, Lotufo RA. User-Steered Image Segmentation Paradigms: Live Wire and Live Lane. Graphical Models Image Process 1998; 60: 233–260.

Fiez JA, Damasio H, Grabowski TJ. Lesion Segmentation and Manual Warping to a Reference Brain: Intra-and Interobserver Reliability. Hum.Brain Mapp. 2000; 9: 192–211.

Filippi M, Horsfield MA, Bressi A, Martinelli V, Baratti C, Reganati P, et al. Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis: A comparison of techniques. Brain 1995; 118: 1593.

Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 1999; 9: 195–207.

Fisher RA. Statistical methods for research workers. 14th ed., revised and enlarged. Edinburgh: Oliver and Boyd; 1970.

Fix E, Hodges J. Discriminatory analysis: non-parametric discrimination: Small sample performance. 1951

Forgy EW. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. Biometrics 1965: 768–769.

Frackowiak R. Human brain function. 2nd ed. Amsterdam ;;Boston: Elsevier Academic Press; 2004.

Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, et al. Bayesian decoding of brain images. NeuroImage 2008; 39: 181–205.

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RSJ. Spatial Registration and Normalization of Images. Hum.Brain Mapp. 1995; 3: 165–189.

Fuchs F, Smith SH. Calcium, cross-bridges, and the Frank-Starling relationship. News Physiol. Sci. 2001; 16: 5–10.

Fukunaga K, Hummels DM. Bayes error estimation using Parzen and k-NN procedures. IEEE Trans. Pattern Anal. Mach. Intell. 1987; 9: 634–643.

Fukunage K, Narendra PM. A Branch and Bound Algorithm for Computing k-Nearest Neighbors. IEEE Transactions on Computers 1975; 24: 750–753.

Geman S, Bienenstock E, Doursat R. Neural Networks and the Bias/Variance Dilemma. Neural Computation 1992; 4: 1–58.

Geva S, Jones PS, Crinion JT, Price CJ, Baron J-C, Warburton EA. The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. Brain 2011; 134: 3071–3082.

Gläscher J, Tranel D, Paul LK, Rudrauf D, Rorden C, Hornaday A, et al. Lesion Mapping of Cognitive Abilities Linked to Intelligence. Neuron 2009; 61: 681–691.

Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation: analysis of repeated Barthel index measures. Arch Phys Med Rehabil 1979; 60: 14–17.

Greicius MD, Kimmel DL. Neuroimaging insights into network-based neurodegeneration: Current Opinion in Neurology 2012; 25: 727–734.

Gupta V, Prakash KNB, Nowinski WL. Towards discrimination of infarcts from artifacts in DWI scans. International Journal of Computer Assisted Radiology and Surgery 2008; 2: 385–395.

Harmeling S, Dornhege G, Tax D, Meinecke F, Muller K-R. From outliers to prototypes: Ordering data. Blind Source Separation and Independent Component Analysis - Selected papers from the ICA 2004 meeting, Granada, Spain, Blind Source Separation and Independent Component Analysis 2006; 69: 1608–1618.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 2001; 293: 2425–2430.

Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? Nat. Rev. Neurosci. 2002; 3: 142–151.

Hendrikse J, Petersen ET, van Laar PJ, Golay X. Cerebral border zones between distal end branches of intracranial arteries: MR imaging. Radiology 2008; 246: 572–580.

Hevia-Montiel N, Jimenez-Alaniz JR, Medina-Banuelos V, Yanez-Suarez O, Rosso C, Samson Y, et al. Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE. 2007. p. 2102–2105.

Holm A, Pedersen SS, Nexoe J, Obel N, Nielsen LP, Koldkjaer O, et al. Procalcitonin versus C-reactive protein for predicting pneumonia in adults with lower respiratory tract infection in primary care. Br J Gen Pract 2007; 57: 555–560.

Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. A practical guide to support vector classification 2003

Husain M, Kennard C. Visual neglect associated with frontal lobe infarction. J Neurol 1996; 243: 652–657.

Hutchison RM, Gallivan JP, Culham JC, Gati JS, Menon RS, Everling S. Functional connectivity of the frontal eye fields in humans and macaque monkeys investigated with resting-state fMRI. J. Neurophysiol. 2012; 107: 2463–2474.

Johnston KC, Connors AFJ, Wagner DP, Knaus WA, Wang X, Haley ECJ. A predictive risk model for outcomes of ischemic stroke. Stroke 2000; 31: 448–455.

Johnston KC, Wagner DP, Wang X-Q, Newman GC, Thijs V, Sen S, et al. Validation of an acute ischemic stroke model: does diffusion-weighted imaging lesion volume offer a clinically significant improvement in prediction of outcome? Stroke 2007; 38: 1820–1825.

Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002; 24: 881–892.

Karnath HO, Ferber S, Himmelbach M. Spatial awareness is a function of the temporal not the posterior parietal lobe. Nature 2001; 411: 950–953.

Karnath HO, Fruhmann BM, Kuker W, Rorden C. The Anatomy of Spatial Neglect based on Voxelwise Statistical Analysis: A Study of 140 Patients. Cerebral Cortex 2004; 14: 1164–1172.

Keerthi SS, Lin CJ. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel. Neural Comput. 2003; 15: 1667–1689.

Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 2009; 46: 786–802.

Kohavi R. A study of cross validation and bootstrap for accuracy estimation and model selection. [S.l.]: International joint conference on artificial intelligence; 1995.

König IR, Ziegler A, Bluhmki E, Hacke W, Bath PMW, Sacco RL, et al. Predicting long-term outcome after acute ischemic stroke: a simple index works in patients from controlled clinical trials. Stroke 2008; 39: 1821–1826.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U.S.A. 2006; 103: 3863–3868.

Kwakkel G, Veerbeek JM, Harmeling-van der Wel B, van Wegen E, Kollen BJ. Diagnostic accuracy of the Barthel Index for measuring activities of daily living outcome after ischemic hemispheric stroke: does early poststroke timing of assessment matter? Stroke 2011; 42: 342–346.

Van Laar PJ, van der Grond J, Hendrikse J. Brain perfusion territory imaging: methods and clinical applications of selective arterial spin-labeling MR imaging. Radiology 2008; 246: 354–364.

Lao Z, Shen D, Liu D, Jawad AF, Melhem ER, Launer LJ, et al. Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. Acad Radiol 2008; 15: 300–313.

Van Leemput K, Maes F, Vandermeulen D, Colchester A, Suetens P. Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans.Med.Imaging 2001; 20: 677–688.

Lin HT, Lin CJ. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Taipei: Department of Computer Science and Information Engineering, National Taiwan University 2003

Liu C, McNeil JE, Greenwood R. Rehabilitation outcomes after brain injury: disability measures or goal achievement? Clin Rehabil 2004; 18: 398–404.

Logothetis NK. The underpinnings of the BOLD functional magnetic resonance imaging signal. J. Neurosci. 2003; 23: 3963–3971.

Lövblad KO, Baird AE, Schlaug G, Benfield A, Siewert B, Voetsch B, et al. Ischemic lesion volumes in acute stroke by diffusion-weighted magnetic resonance imaging correlate with clinical outcome. Ann. Neurol. 1997; 42: 164–170.

Lyden P, Claesson L, Havstad S, Ashwood T, Lu M. Factor analysis of the National Institutes of Health Stroke Scale in patients with large strokes. Arch. Neurol. 2004; 61: 1677–1680.

Van der Maaten L, Hinton G. Visualizing Data using t-SNE. The Journal of Machine Learning Research 2008; 9: 2579–2633.

MacKay D. Information Theory, Inference, and Learning Algorithms. Cambridge University Press; 2003.

MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Univ. of Calif. Press; 1967.

Mah Y-H, Jager R, Kennard C, Husain M, Nachev P. A new method for automated high-dimensional lesion segmentation evaluated in vascular injury and applied to the human occipital lobe [Internet]. Cortex 2012[cited 2013 Jan 5] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0010945212003437

Mahalanobis P. Normalization of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributinos. Sankhya 1937; 1: 35–45.

Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. Md State Med J 1965; 14: 61–65.

Matthews PM, Jezzard P. Functional magnetic resonance imaging. J. Neurol. Neurosurg. Psychiatr. 2004; 75: 6–12.

Menon, Kim. Spatial and temporal limits in cognitive neuroimaging with fMRI. Trends Cogn. Sci. (Regul. Ed.) 1999; 3: 207–216.

Mesulam MM. A cortical network for directed attention and unilateral neglect. Ann. Neurol. 1981; 10: 309–325.

Milhorn HTJ, Benton R, Ross R, Guyton AC. A mathematical model of the human respiratory control system. Biophys. J. 1965; 5: 27–46.

Mort DJ, Malhotra P, Mannan SK, Rorden C, Pambakian A, Kennard C, et al. The anatomy of visual neglect. Brain 2003; 126: 1986–1997.

Mourao-Miranda J, Reinders AATS, Rocha-Rego V, Lappin J, Rondina J, Morgan C, et al. Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. Psychological Medicine 2011; 42: 1037–1047.

Nachev P, Coulthard E, Jager HR, Kennard C, Husain M. Enantiomorphic normalization of focally lesioned brains. NeuroImage 2008; 39: 1215–1226.

Nachev P. Barking up the wrong tree: the problem with mass univariate analysis in lesion-function brain mapping [Internet]. 2008Available from: http://mta.hu/fileadmin/2008/04/Budapest_Workshop_schedule.pdf

Naess H, Waje-Andreassen U, Thomassen L, Myhr K-M. High incidence of infarction in the left cerebral hemisphere among young adults. J Stroke Cerebrovasc Dis 2006; 15: 241–244.

Nakao S, Takata S, Uemura H, Kashihara M, Osawa T, Komatsu K, et al. Relationship between Barthel Index scores during the acute phase of rehabilitation and subsequent ADL in stroke patients. J. Med. Invest. 2010; 57: 81–88.

Nazzal ME, Saadah MA, Saadah LM, Trebinjac SM. Acute ischemic stroke: relationship of brain lesion location & functional outcome. Disabil Rehabil 2009; 31: 1501–1506.

Neal R. Probabilistic Inference Using Markov Chain Monte Carlo Methods, Technical Report CRG-TR-93-1, Dept. of Computer Science [Internet]. 1993Available from: http://www.cs.toronto.edu/~radford/ftp/review.pdf

Neal R. Flexible Bayesian Modeling and Markov Chain Sampling. 2004.

Nolf E, Voet T, Jacobs F, Dierckx R, Lemahieu I. XMedCon : an Open-source Medical Image Conversion Toolkit. European Journal of Nuclear Medicine and Molecular Imaging 2003; 30: S246–S246.

Nyul LG, Udupa JK. On Standardizing the MR Image Intensity Scale. Magnetic resonance imaging 1999; 42

Özsu MT. Encyclopedia of database systems. New York; London: Springer; 2009.

Pantazis D, Joshi A, Jiang J, Shattuck DW, Bernstein LE, Damasio H, et al. Comparison of landmark-based and automatic methods for cortical surface registration. Neuroimage 2010; 49: 2479–2493.

Paus T, Jech R, Thompson CJ, Comeau R, Peters T, Evans AC. Transcranial magnetic stimulation during positron emission tomography: a new method for studying connectivity of the human cerebral cortex. J. Neurosci. 1997; 17: 3178–3184.

Plant C, Teipel SJ, Oswald A, Böhm C, Meindl T, Mourao-Miranda J, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 2010; 50: 162–174.

Poldrack R. Handbook of functional MRI data analysis. New York: Cambridge University Press; 2011.

Prastawa M, Bullitt E, Ho S, Gerig G. A brain tumor segmentation framework based on outlier detection. Med Image Anal 2004; 8: 275–283.

Rieck K, Laskov P. Detecting unknown network attacks using language models. Lecture Notes in Computer Science 2006; 4064: 74–90.

Rieck K, Laskov P. Language models for detection of unknown attacks in network traffic. J Comput Virol 2007; 2: 243–256.

Ripollés P, Marco-Pallarés J, de Diego-Balaguer R, Miró J, Falip M, Juncadella M, et al. Analysis of automated methods for spatial normalization of lesioned brains. Neuroimage 2012; 60: 1296–1306.

Rivers CS, Wardlaw JM, Armitage PA, Bastin ME, Hand PJ, Dennis MS. Acute Ischemic Stroke Lesion Measurement on Diffusion-weighted Imaging–Important Considerations in Designing Acute Stroke Trials With Magnetic Resonance Imaging. Journal of Stroke and Cerebrovascular Diseases 2007; 16: 64–70.

Rollnik JD. Barthel index as a length of stay predictor in neurological rehabilitation. Rehabilitation (Stuttg) 2009; 48: 91–94.

Rorden C, Brett M. Stereotaxic display of brain lesions. Behavioural Neurology 2001; 12: 191–200.

Rorden C, Karnath HO. Using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat. Rev. Neurosci 2004; 5: 813–819.

Rudiak D, Marg E. Finding the depth of magnetic brain stimulation: a re-evaluation. Electroencephalogr Clin Neurophysiol 1994; 93: 358–371.

Ruff CC, Driver J, Bestmann S. Combining TMS and fMRI: from 'virtual lesions' to functional-network accounts of cognition. Cortex 2009; 45: 1043–1049.

Rychetsky M. Algorithms and architectures for machine learning based on regularized neural netwoks and support vector approaches. 2001

Sack AT, Linden DEJ. Combining transcranial magnetic stimulation and functional imaging in cognitive brain research: possibilities and limitations. Brain Research Reviews 2003; 43: 41–56.

Saunders DE, Clifton AG, Brown MM. Measurement of infarct size using MRI predicts prognosis in middle cerebral artery infarction. Stroke 1995; 26: 2272–2276.

Saur D, Ronneberger O, Kümmerer D, Mader I, Weiller C, Klöppel S. Early functional magnetic resonance imaging activations predict language outcome after stroke. Brain 2010; 133: 1252–1264.

Savoiardo M. The vascular territories of the carotid and vertebrobasilar systems. Diagrams based on CT studies of infarcts. Ital J Neurol Sci 1986; 7: 405–409.

Schiemanck SK, Kwakkel G, Post MWM, Kappelle LJ, Prevo AJH. Predicting long-term independency in activities of daily living after middle cerebral artery stroke: does information from MRI have added predictive value compared with clinical information? Stroke 2006; 37: 1050–1054.

Scoville WB, Milner B. Loss of Recent Memory After Bilateral Hippocampal Lesions. J Neurol Neurosurg Psychiatry 1957; 20: 11–21.

Seghier ML, Ramlackhansingh A, Crinion J, Leff AP, Price CJ. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. Neuroimage 2008; 41: 1253–1266.

Shaywitz BA, Pugh KR, Todd Constable R, Shaywitz SE, Bronen RA, Fulbright RK, et al. Localization of semantic processing using functional magnetic resonance imaging. Human Brain Mapping 1994; 2: 149–158.

Shelton FN, Reding MJ. Effect of lesion location on upper limb motor recovery after stroke. Stroke 2001; 32: 107–112.

Shen S, Szameitat AJ, Sterr A. VBM lesion detection depends on the normalization template: a study using simulated atrophy. Magn.Reson. Imaging 2007; 25: 1385–1396.

Shen S, Szameitat AJ, Sterr A. Detection of Infarct Lesions From Single MRI Modality Using Inconsistency Between Voxel Intensity and Spatial Location: A 3-D Automatic Approach. Information Technology in Biomedicine, IEEE Transactions on 2008; 12: 532–540.

Shen S, Szameitat AJ, Sterr A. An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps. Magn Reson Imaging 2010; 28: 245–254.

Smith DV, Clithero JA, Rorden C, Karnath H-O. Decoding the anatomical network of spatial attention. Proceedings of the National Academy of Sciences 2013; 110: 1518–1523.

Stanton GB, Bruce CJ, Goldberg ME. Topography of projections to posterior cortical areas from the macaque frontal eye fields. J. Comp. Neurol. 1995; 353: 291–305.

Stone SP, Allder SJ, Gladman JR. Predicting outcome in acute stroke. Br. Med. Bull. 2000; 56: 486–494.

Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. Stroke 1999; 30: 1538–1541.

Tahmasebi AM, Abolmaesumi P, Zheng ZZ, Munhall KG, Johnsrude IS. Reducing inter-subject anatomical variation: effect of normalization method on sensitivity of functional magnetic resonance imaging data analysis in auditory cortex and the superior temporal region. Neuroimage 2009; 47: 1522–1531.

Tan IL, Schijndel RA, Walderveen MAA, Quist M, Bos R, Pouwels PJW, et al. Magnetic resonance image registration in multiple sclerosis: Comparison with repositioning error and observer-based variability. Journal of Magnetic Resonance Imaging 2002; 15: 505–510.

Tax D, Duin R. Data domain description using support vectors. Bruges, Belgium: 1999. p. 251–256.

Tenenbaum JB. A Global Geometric Framework for Nonlinear Dimensionality Reduction. The Journal of Machine Learning Research 2000; 9: 2319–2323.

The FiL Methods group. SPM [Internet]. The FiL Methods group; Available from: http://www.fil.ion.ucl.ac.uk/spm/

Thiebaut de Schotten M, Dell'Acqua F, Forkel SJ, Simmons A, Vergani F, Murphy DGM, et al. A lateralized brain network for visuospatial attention. Nat. Neurosci. 2011; 14: 1245–1246.

Thijs VN, Lansberg MG, Beaulieu C, Marks MP, Moseley ME, Albers GW. Is early ischemic lesion volume on diffusion-weighted imaging an independent predictor of stroke outcome? A multivariable analysis. Stroke 2000; 31: 2597–2602.

Thompson PM, Schwartz C, Lin RT, Khan AA, Toga AW. Three-dimensional statistical analysis of sulcal variability in the human brain. J. Neurosci. 1996; 16: 4261–4274.

Tipping M. Bayesian Inference: An Introduction to Principles and Practice in Machine Learning. Advanced Lectures on Machine Learning 2004; 3176: 41–62.

Tipping ME. Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research 2001; 1: 211–244.

Tong DC, Yenari MA, Albers GW, O'Brien M, Marks MP, Moseley ME. Correlation of perfusion- and diffusion-weighted MRI with NIHSS score in acute (<6.5 hour) ischemic stroke. Neurology 1998; 50: 864–870.

Tosun D, Rettmann ME, Prince JL. Mapping techniques for aligning sulci across multiple brains. Med Image Anal 2004; 8: 295–309.

Utz KS, Dimova V, Oppenländer K, Kerkhoff G. Electrified minds: transcranial direct current stimulation (tDCS) and galvanic vestibular stimulation (GVS) as methods of non-invasive brain stimulation in neuropsychology--a review of current data and future implications. Neuropsychologia 2010; 48: 2789–2810.

Uyttenboogaart M, Stewart RE, Vroomen PCAJ, De Keyser J, Luijckx G-J. Optimizing cutoff scores for the Barthel index and the modified Rankin scale for defining outcome in acute stroke trials. Stroke 2005; 36: 1984–1987.

Vallar G, Perani D. The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. Neuropsychologia 1986; 24: 609–622.

Vapnik V. Estimation of Dependences Based on Emperical Data. 1982.

Verdon V, Schwartz S, Lovblad K-O, Hauert C-A, Vuilleumier P. Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping. Brain 2009; 133: 880–894.

Vuilleumier P, Hester D, Assal G, Regli F. Unilateral spatial neglect recovery after sequential strokes. Neurology 1996; 46: 184–189.

Walsh V, Cowey A. Transcranial magnetic stimulation and cognitive neuroscience. Nat. Rev. Neurosci. 2000; 1: 73–79.

Warach S, Dashe JF, Edelman RR. Clinical outcome in ischemic stroke predicted by early diffusion-weighted and perfusion magnetic resonance imaging: a preliminary analysis. J. Cereb. Blood Flow Metab. 1996; 16: 53–59.

Warach S, Gaa J, Siewert B, Wielopolski P, Edelman RR. Acute human stroke studied by whole brain echo planar diffusion-weighted magnetic resonance imaging. Ann. Neurol. 1995; 37: 231–241.

Watkins KE, Paus T, Lerch JP, Zijdenbos A, Collins DL, Neelin P, et al. Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. Cereb. Cortex 2001; 11: 868–877.

Weiller C, Isensee C, Rijntjes M, Huber W, Müller S, Bier D, et al. Recovery from Wernicke's aphasia: a positron emission tomographic study. Ann. Neurol. 1995; 37: 723–732.

Wernicke C. {Der aphasische Symptomencomplex}. Cohn und Weigart; 1874.

Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA Stroke Study Group. Stroke 1999; 30: 2355–2359.

Yassa MA, Stark CEL. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. Neuroimage 2009; 44: 319–327.

Zangen A, Roth Y, Voller B, Hallett M. Transcranial magnetic stimulation of deep brain regions: evidence for efficacy of the H-coil. Clin Neurophysiol 2005; 116: 775–779.

Zhu LL, Lindenberg R, Alexander MP, Schlaug G. Lesion Load of the Corticospinal Tract Predicts Motor Impairment in Chronic Stroke. Stroke 2010; 41: 910–915.

Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. Medical Imaging, IEEE Transactions on 1994; 13: 716–724.

MATLAB. Natick, MA 2000: The MathsWorks Inc;

Paraview [Internet]. Kitware; Available from: http://www.paraview.org/